Efficient Multiband Algorithms for Blind Source Separation

Salah Al-Din Ibrahim Badran

BSc, MPhil

A thesis submitted in fulfilment of

the requirement for the degree of

Doctor of Philosophy

At

De Montfort University

March 2016

Faculty of Technology

Department of Computing & Electrical Engineering

Supervisors: **Dr. Samad Ahmadi**

Dr. Ismail Shahin (University of Sharjah)

Dr. Pooneh Zadeh

Abstract

The problem of blind source separation (BSS) refers to recovering original audio signals, called source signals, from the mixed signals, called convolutive mixtures or observation signals, in a reverberant environment. The mixture is a function of a sequence of original speech signals mixed in a reverberant room or ambient. The objective is to separate mixed signals to obtain the original signals without degradation and without prior information of the features of the sources, such as the locations, the spectral nature or the mixing method. The strategy used to achieve this objective is to use multiband schemes. Multiband schemes work at a lower sampling rate, have less computational cost and a quicker convergence than the full-band approach. Our motivation is the competitive results of unequal-passbands scheme applications, in terms of the convergence speed and computational complexity. The objective of this research is to improve unequal-passbands schemes by improving the speed of convergence and reducing the computational cost. The first proposed work is a novel maximally decimated unequal-passbands scheme. This scheme is applied to system identification. This scheme uses an arbitrary number of bands; multiple bands make it work at a reduced sampling rate, and low computational cost. An adaptation approach is derived with a normalised adaptation step that improved the convergence speed.). The performance of the proposed scheme was measured in different ways. First, the mean square errors of various bands are measured and the results are compared to a maximally decimated equal-passbands scheme, which is currently the best performing method. The results show that the proposed scheme has a faster convergence rate than the maximally decimated equal-passbands scheme. Second, when the scheme is tested for white and coloured inputs using a low number of bands, it does not yield good results; but when the number of bands is increased, the speed of convergence is enhanced. Third, the scheme is tested for quick changes in the system identification. It is shown that the performance of the

proposed scheme is similar to that of the equal-passbands scheme. Fourth, the scheme is also tested in a stationary state for both colour and white signals. The experimental results confirm the theoretical work. The maximally decimated unequal-passbands scheme has a lower computation cost than maximally decimated equal-passbands scheme and is the best scheme for the system identification scenario, as is proven in the experiments. For more challenging scenarios, an unequal-passbands scheme with oversampled decimation is proposed for blind source separation application; the greater number of bands, the more efficient the separation. The superior performance of this scheme is proved by: first studying the simulation time and the signal-to-interference ratio of the novel normalisation. The results are compared to the currently best performing method. Second, an experimental comparison is made between the proposed multiband scheme and the conventional full-band scheme. The results show that the convergence speed and the signal-to-interference ratio of the proposed scheme are higher than that of the full-band scheme, and the computation cost is lower than that of the full-band scheme.

Acknowledgments

I would like to express my warmest thanks to my supervisors who were given me their supports and encouragement. A special thanks to Dr. Ismail Shahin from the University of Sharjah for his continuous help.

I am grateful to my sister Dr. H. Badawi for her invaluable help. The simplest words being the strongest, I address all my passions to my family and especially to my mom and father who made me understand that life is not made those problems that could be solved through mathematical formulas and algorithms. Despite being away for (too) many years, their intelligence, their confidence, their tenderness, their loves carry me and guide me all days. Thank you for making me what I am today. Is this a good place for to say that kind of things? I do not know any bad. I love you. A thought to end these thanks for you who did not see the outcome of my work but I know you would have been very proud of your son!

Finally, big thanks and love to my wife Wafa who stood beside me and encourage me all the time. I would like to thank my children: Sarah, Yasmin, Afnan, and Ahmed for their patience and understanding for all the time that I spent away from them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AP	Affine Projection
bdiag	block- diagonal
Boff	block-off-diagonal
BSS	Blind Source Separation
CMFB	Cosine Modulated Filter Bank
DCT	Discrete Cosine Transform
DOA	Direction Of Arrive
EEG	Electro-Encephalogram
EP	Equal-Passbands
EPS	Equal-passbands Scheme
FB	Full-band
FD	Frequency Domain
FIR	Finite Impulse Response
HOS	Higher Order Statistics
HPF	High-Pass Filter
IBM	Ideal Binary Mask
ICA	Independent Component Analysis
IDFT	Inverse Discrete Fourier Transform
IIR	Infinite Impulse Response
ITU	International Telecommunication Union
KMC	Kurtosis Maximization Criterion

LMS	Least-Mean-Square
LPF	Low-Pass Filter
LTI	Linear Time Invariant
MB	Multiband/Multiple bands
MD	Maximally Decimated
MDP	Minimal Distortion Principle
ML	Maximum Likelihood
MOS	Mean Opinion Score
MSE	Mean Square Error
ND	Non-stationary Decorrelation
NLMS	Normalised LMS
NMD	Non-Maximally Decimated
NPR	Near Perfect Reconstruction
PCA	Principal Component Analysis
PDF	Probability Density Function
PESQ	Perceptual Evaluation Signal Quality
PFB	Perceptual Filter Bank
PQMF	Pseudo Quadrature Mirror Filter
PR	Perfect Reconstruction
PSD	Power Spectral Density
QMR	Quadrature Mirror Filter
RASTI	Room Acoustics or Rapid Speech Transmission Index
RLS	Recursive-Least-Square

RT	Reverberation Time
SDR	Signal-to-Distortion Ratio
SIR	Signal-to-Interference Ratio
SiSEC	Signal Separation Evaluation Campaign Community
SNR	Signal-to-noise Ratio
SOBI	Second Order Blind Identification
SOS	Second Order Statistics
SSNR	Segmental SNR
STFD	Spatial Time-Frequency Distribution
STFR	Spatial Time-Frequency Representations
STFT	Short-Time Fourier Transform
STI	Speech Transmission Index
TD	Time Domain
TDD	Time-Delayed Decorrelation
TDSEP	Temporal Decorrelation Source Separation
T-F	Time-Frequency domain
TF-DSS	Time-Frequency-Denoising Source Separation
TF-GEVD	Time-Frequency-Generalised EigenValue Decomposition
UEP	Unequal-Passbands
UEPS	Unequal-passbands Scheme

Chapter 1
Introduction

1.1 **Background and Motivation**

The parameters of the adaptive filter are continuously adjusted by an adaptation algorithm, which aims to optimise some performance metrics. Typically, the definition of the criterion of performance requires a reference input, which is generally ignored, when constructing fixed filters.

Fig. 1.1 presents a general framework for supervised adaptive filtering [**1**], where $x(k)$ and $y(k)$ are the input and output signals, respectively, and the reference signal is represented by $r(k)$. The error signal

$$e(k) = r(k) - y(k), \tag{1.1}$$

is used in the structure of the objective function (also called performance or cost function) which is responsible for updating the adaptive filter coefficients (taps). Different schemes can affect the computational cost (number of mathematical operations per iteration).

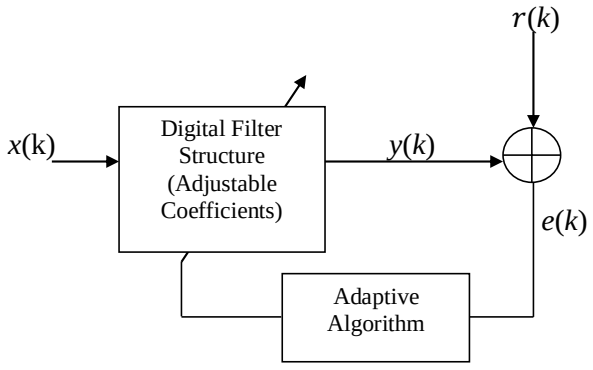


Fig. 1.1 A Supervised Adaptive filter

In recent decades, several algorithms have been developed [**1**]. Examples of these algorithms are: (1) the conventional Least-Mean-Square (LMS) algorithm that has low computational cost, but its drawback is when the input signals are strongly correlated (coloured signals) (2) the normalised LMS (NLMS) algorithm, which offers a higher computational cost than the conventional LMS algorithm in the frequency domain (FD), for which high-order adaptive filters have reduced the computational cost and have fast convergence to colour signals in relation to the conventional LMS, and (3) Recursive-Least-Square (RLS) algorithm, which has a high convergence speed, but a high computational cost and possible numerical instability. Fig. 1.2 shows the basic adaptive filter structure with multiple bands. Essentially, the input and desired signals are partitioned into L adjacent bands using an analysis filter bank, then in each band signal is subsampled and the LMS algorithm is applied using the error of each band to adapt its adaptive filter. From there, the signals are interpolated and combined by a synthesis filter bank, generating the error signal of the scheme $e(k)$, with which the Mean Square Error (MSE) of the scheme will be calculated. In this case, the finite impulse response will be a composition of the non-trivial impulse responses of filters. This subsampling reduces the computational complexity. On the other hand, the reduction implies an improvement in the condition of the correlation matrix [**1**]. Another factor that contributes to the acceleration of the rate of convergence is the normalisation step of adapting the energy of each band signal input of the respective adaptive filter.

The blind source separation is the second part of our work; the basic structure is illustrated in Fig. 1.3 [**1**]. This problem has been investigated in recent decades. The signal $x_m(k)$ is captured by M sensors (microphones) as a result of a linear convolutive mixture, see Section 1.2.2.3, of N noise sources and $U_n(k)$ sound sources. The output signals that represent the estimates of the source signals are obtained without any knowledge of their positions, their spectral contents, and/or the mixing system; hence, why it was coined “blind source separation”.

There are numerous applications of signal separation, such as audio systems [**2**], remote sensing [**3**], analysis of seismic signals [**4**], image enhancement [**5**] and digital communications [**6**].

A model of multiple inputs and multiple outputs is used with N inputs and M outputs, and each signal $x_m(k)$ represents the sum of sources convolved with the impulse responses from each set of N sources until the m-th microphone. This mixing system is modelled as FIR filters whose orders (typically a few thousand) depend on the reverberation time of the environment [**7**].

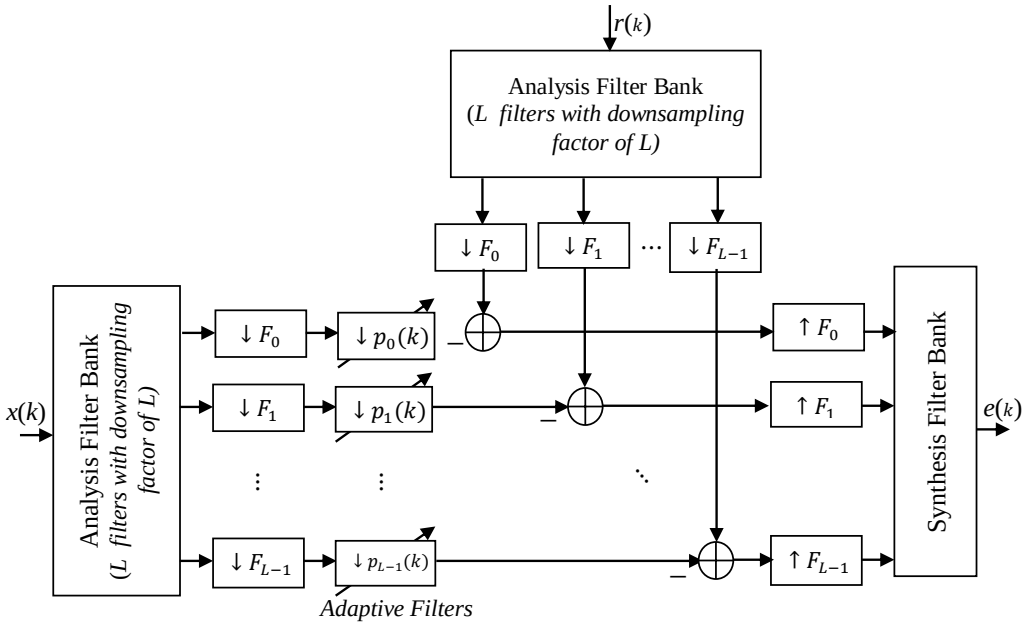


Fig. 1.2 An example of adaptive filtering using multiple bands.

The adaptive learning algorithm can be the gradient descent type [**1**], but the objective function to be minimised takes into account the mutual information between the signals $x_m(k)$, so that the coefficients of the separation system are adapted so that the signals $y_r(k)$ become mutually independent and satisfy the estimates of the sources.

The speech signals of multiple speakers in a room are mixed with other audio sources, such as music [**8**], and noises [**9**]. Different methods have proposed to separate the mixed signals for different applications, such as noise cancellation and speech processing for people with hearing difficulties [**10**]. Many researchers have suggested different statistical solutions for blind separation applications in audio signal processing and cognitive psychology [**11**].

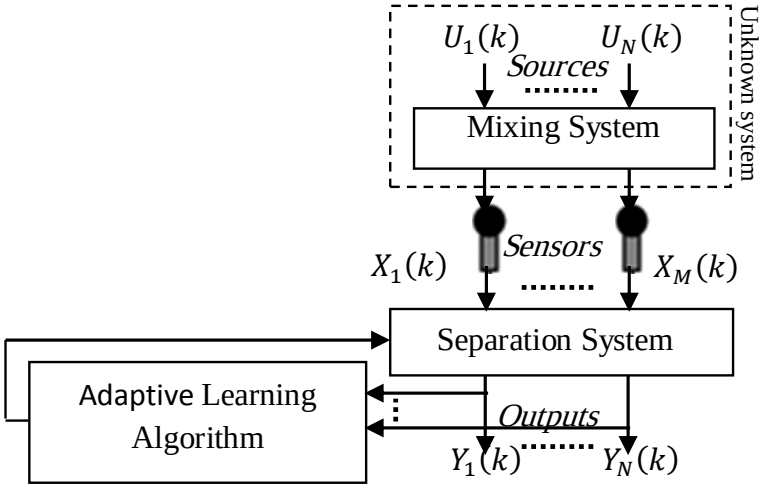


Fig. 1.3 The basic structure of Blind Source Separation System

Nowadays, speech recognition technology is desirable in speaker identification applications [**12**]. Though, the recognition worsens rapidly when many people talk at the same time or background noise is added to the speech signals (e.g., air-conditioning, engine noises of machinery, noise, etc.). This challenge motivates researchers to find a way to create new devices and methods that extract the original speech signals from the undesired signals. This scenario is often referred to as the "cocktail party problem" [**13**], which seeks to detect an individual speech among mixtures captured by microphones.

The motivation for this study are good performing results, which showed that the UEPS has a superior performance over the EPS [**14**] for applications that need a huge number of parameters, such as system identification and convolutive mixtures BSS.

1.2 Principle of Blind Source Separation

The principle of BSS can be stated as follows: it is required to reconstruct the *N* sources signals that are received by *M* mixtures. The processing is then blind, i.e., it has no prior knowledge of the mixing system and the sources are unobservable. It is necessary to have an additional assumption; otherwise, the blind source separation appears as an unsolvable problem. This is why most of the BSS techniques assume that

the sources are independent. Initially, it was modelled based on a single product, i.e., the sensors receive at each instant a linear combination of the source signals. Later, modelling close to reality was introduced. One of these models considers transmission channel as a system based on a filtering operation. In other words, the captured signals are linearly dependent on both the source and their delayed versions. The mixing system is said to be a linear convolutive system.

Different approaches have offered different solutions to the BSS problem, some of them were stated above. One approach is to use the Independent Component Analysis (ICA) technique, formalized by Comon in [**15**].

1.2.1 *Nature of the Mixture*

The nature of the mixture represents the assumptions made about the acoustic environment in which recording is made. In particular, the recordings that occur in a natural environment are affected by multiple echoes that result in multiple directions for each source. If the recording is done in an anechoic room, where it can be assumed that there is no echo, however the distances from a source at each microphone are different, the signals of the sources do not arrive at the same time to the different microphones.

1.2.2 *Mixture and Instantaneous Linear Model*

1.2.2.1 *Instantaneous Mixture*

This is the simplest case that is usually used for theoretical derivation as it neglects the reverberation time (RT) in the room. The source separation is a signal processing technique in which multiple sensors (microphones) are expressed as a sequence of a M-dimensional observations $x(t)$. It is modelled as

$$x(t) = Au(t) + n(t) \tag{1.2}$$

The matrix A is of $M \times N$ dimension, and has the coefficients of the linear time invariant (LTI) mixing system. The variable $u(t)$ is an N-source vector whose components

are independent for each t and $n(t)$ represents any additive noise, and $t = 1,\ldots T$, is the time.

The assumptions are therefore very similar to those of the ICA. However, the existence of a possible time dependency between successive samples makes it possible to exploit the statistical independence between the source signals other than in a non-Gaussian model. This description is difficult because there is no uniformly efficient algorithm for several reasons. First, there are many ways of expressing independence. If we choose to rely on a possible temporal (or spatial) structure of the observations as suggested by equation (1.2), it is possible to be satisfied with Gaussian models. We can then use the spectral diversity (or non-stationarity) of the observations through the formation of spectral covariance matrices.

Finally, it can also be effective to extract the components sequentially for each component by using fixed-point algorithms such as Fast-ICA.

1.2.2.2 *Mathematical model of instantaneous mixture*

The observed signal is $x_i(t)$, superposition of signals $u_j(t)$. The linear instantaneous mixture model is expressed as follows:

$$x_i(t) = \sum_{j=1}^{N} a_{ij} u_j(t) + n_i(t) \tag{1.3}$$

with $i \in [1,2,3,\cdots,M]$. Equation (1.3) gives the vector shape of this expression, at time t. The instantaneous mixture model therefore reflects a spatial relationship between sources and sensors at every moment.

1.2.2.3 *convolutive mixtures model*

Equation (1.3) describes an instantaneous mixture, where the reverberation is neglected. In the real world, the original signals are recorded in a reverberant room and

they are a convolutive mixture [**16**]. Ignoring the noise $n(t)$, the problem may be formulated as follows:

$$x(t) = A*u(t) \tag{1.4}$$

where the asterisk denotes convolution operation. The inverse matrix W, that we aim to find, is included in the following equation:

$$\tilde{u}(t) = W*x(t) \tag{1.5}$$

where W is the separation matrix, and the components of $\tilde{u}(t)$ as independent as possible. The calculation of the convolutive mixture in the time domain (TD) becomes complex and slow because the convolution operation is computationally expensive in the time domain. If the equation is transformed into a FD, the convolution operation will be changed into multiplication. This means that it is possible to apply the methods of instantaneous mixtures at every frequency.

$$X = AU \tag{1.6}$$

$$\tilde{U} = WX \tag{1.7}$$

The first work on the separation of the sources from the convolutive mixtures was done in the time domain [**17**]. Inspired by the methods of blind deconvolution, Torkkola [**18**] modelled the separation process as an FIR filter which must be estimated from the filter coefficients ω_{rc}:

$$\tilde{u}(t) = \sum_{k=0}^{K} W(k)x(t-k) \tag{1.8}$$

A method of maximising the mutual information is suggested by Bell and Senjnowski [**19**]. Lee [**20**] has modelled the separation procedure, as an IIR filter, assuming that the recording environment is at minimum phase, which is not always valid. Mukai et al. in [**21**] and [**22**] have studied the separation of convolutive mixtures problems. They used the STFT:

$$|U_x(t,f)| = A(f)U_u(t,f) \tag{1.9}$$

where *A(f)* is the Fourier transform of the convolution filter, $U_u(t,f)$ is the sources in time-frequency (T-F), and $U_x(t,f)$ is the vector of time-frequency (T-F) of the mixture. By applying *W(f)*, the separation matrix, for each frequency

$$U_{\tilde{u}}(t,f) = W(f)U_x(t,f)\,. \tag{1.10}$$

Smaragdis [**23**] has also conducted research in the same field to transform the convolution multiplication. The problem is defined as an estimate of the separation matrix for each frequency frame which may be a solution in the sense of maximum likelihood. Unfortunately, solving the problem independently for each frequency frame produces permutation problems. The permutation means that the first signal at the output of the BSS process is assured to be the first input source [**24**]. To resolve this issue Smaragdis proposed an adaptive scheme to apply coupling frequencies for neighbouring frames.

1.2.3 *Determination of the Mixture*

There are two kinds of mixtures (microphones), linear and nonlinear. The linear mixture can be further classified into instantaneous mixtures and convolutive mixtures, and time-varying and time-invariant [**25**]. These types can be classified further as:

1. Determined case: $M = N$ (same number of sensors and sources);

2. Overdetermined case: $M > N$ (more sensors than sources);

3. Underdetermined case: $M < N$ (fewer sensors than sources).

1.2.4 *Level of a information on the Sources and Mixture*

Estimating the mixture will be easier, if one knows the number of sources, physical location of sources and sensors, or the characteristics and the relative position of the sensors. Similarly a certain number of a prior knowledge on sources, for example,

spectral nature of information on the sources, can solve the problem of blind separation of sources more efficiently.

1.3 List of Contributions

An unequal-passbands (non-uniform) scheme (UEPS) is used in our two proposals. A tree-structured UEPS is used and designed from a 2-channel equal-passbands (uniform) filter bank [**14**]. This kind of UEPS is useful for applications that need a large number of parameters such as system identification and convolutive mixtures blind source separation (BSS), which are our proposed works.

1.3.1 *Addressing the problem:*

The first proposal is a maximally decimated UEPS with multiple bands. This scheme is suitable for modelling any FIR system with an arbitrary number of bands and minimised aliasing. This modelling will improve the convergence speed and lower the adaptive algorithm computational cost. In order to achieve these improvements, a new normalisation approach for the adaptive algorithm is proposed and a significant reduction of the mean square errors between the bands is obtained. This reduction is done by using a reduced decimation rate that can be applied by decomposing the frequency of the input signal.

The second proposal is a blind source separation approach for reverberant rooms (convolutive mixtures) that uses NMD_UEPS filter bank scheme with real-coefficients, to extract the original speech with a better convergence speed and lower computational complexity. To achieve this goal, this proposed structure uses a novel normalisation approach for the adaptation algorithm. Multiple bands are used in the separation system. The rationale for using multiple bands is to enhance the convergence speed and decrease the computational cost as it is compared to the full-band algorithm.

1.3.2 *Importance for methodology development and applications:*

The first proposed work is extended from [**26**], as the new structure is capable of working with an arbitrary number of bands, whereas the old one is restricted to a specific number of bands. To achieve our goal, an UEPS is designed from an adaptive multiband structure that uses an analysis filter bank and non-zero adaptive filters [**27**], see Section 4.7.

The maximally decimated UEPS can be proficiently implemented by carrying out the filtering at the multiple bands that operate at the highest sampling rate employing block processing. An extra reduction can be achieved in the complexity by using the octave-band tree-structured filter banks and employing fast Fourier transforms (FFT), as mentioned in [**28**]. Section 5.7 presents how the convergence rate of the proposed method is enhanced for coloured signals at the input with regard to the full-band LMS method. An adaptation algorithm is derived.

The adaptation convergence is achieved by breaking down the input signal at low frequencies; results in a reduced power spectral density (PSD) ratio. This structure is mostly used for applications that need a large number of adaptive coefficients, for example, Acoustic Echo Cancellation, BSS, and Equalizer filter.

In the second proposed structure, a multiband parameters in the separation stage are set individually using an adaptation method in the time domain [**28**] that uses a novel gradient normalisation approach that leads to a rapid convergence rate and less computational complexity in comparison with the old normalisation approach. The goal of using multiple bands in our structure is to deal with the low frequencies that exist in the speech signal, as the input frequency will be divided by these bands into smaller frequencies. This will reduce the sampling rates, which results in an increased convergence rate and reduced complexity. In addition to that, low frequencies allow us to deal with the large energy that exists in the speech signal. The proposed algorithm has an additional benefit that utilises real filters coefficients that are used in DSP applications; it is also used in the BSS applications.

1.3.3 *The elements of the problem:*

For both proposed methods, the elements of the problem are:

1. increasing the convergence speed,
2. decreasing the computational complexity of the adaptive algorithms, especially for implementations that need a large number of coefficients.

1.3.4 *First contribution*

A maximally decimated UEPS with multiple bands is proposed. This structure is able to model any FIR system employing an arbitrary number of bands. The aliasing effects between bands are eliminated. Using filters with UEPS will break down the input signal frequency and each adaptive filter will operate at different rates. Different rates will increase the quality of the proposed structure and allow us to deal with different normalisation step sizes. Errors that are distributed over multiple bands will make the adjustment of the coefficients easier in the update equation. Our proposed structure enhances the convergence speed and minimises the computational complexity, especially for systems that need a large number of coefficients. Our method is used in system identification. More details are presented in Section 5.4.

1.3.4.1 ***Relation to existing methods:***

Bruno et al. [**29**] studied the comparison of the performance of using multiple bands (MB) and full-band (FB) methods in system identification. The comparison is made on maximal decimation and oversampling versions in time-domain. The results showed that the convergence rate of the multiband (MB) method is in agreement with the full-band method. The computational cost of the multiband method is smaller than the full-band method. In our proposal, we used the LMS algorithm with a maximally decimated UEPS, with multiple bands, and a normalised adaptation step. Section 5.4 explains in detail the performance of the proposed method.

Gilloire and Vetterli [**30**] used the same maximally decimated scheme of Bruno et al. [**29**] but with only two bands. However, their analysis was restricted to a limitation affected by the aliasing of the filter bank. In contrast to our proposed method, we used different bands and found out that the optimal convergence was with four bands. The convergence rate of our suggested method is better than that of Gilloire and Vetterli's method.

Yamada et al. [**31**] were the first to use polyphase filter banks to lower the complexity of the structure. This method worked without degradation of the aliasing, but it converges very slowly with white and colour input signals. Our proposed method employed the same polyphase technique that helps in designing a perfect reconstruction filter bank to avoid reconstruction errors. Section 4.5 gives a detailed explanation with equations and figures of the polyphase technique. This technique, in addition to the new normalisation, helps increase the speed of the convergence rate in comparison to the Yamada structure.

Boroujeny and Wand [**32**] designed analysis and synthesis filters with low delay for implementation in acoustic echo cancellation. To obtain a low delay, the optimum coefficients of the FIR filter are obtained by minimising the energy function that contains the real part of the input signal. The minimum eigenvalues can be achieved when $a^T a = 1$, where a_i are the filter coefficients. On the other hand, in our proposal, we find the optimal coefficients that are extracted by monitoring the coefficient error of the i^{th} band and taking into account the overlapping between the neighbour bands. These results will be formulated within an equation and then used to obtain the expected delay between the bands. Details of our proposed procedure are explained in Section 5.5. We will see that the proposed procedure will lead us to a big matrix where we can see the distribution of the eigenvalue of this matrix that will permit us to predict the performance of the algorithm for a given analysis bank and choose the best number of bands with regard to the convergence rapidity of the adaptation algorithm.

Sridharan [38] presented a new oversampled multiband filtering scheme that is based on a new fast FIR technique; this is achieved by the restriction of non-neighbouring filters. They used Normalised LMS (NLMS) that is derived for the main channels by

reducing the mean squared error and cross filters to remove the aliasing between adjacent channels. The latter filters are derived from the main channel filters. The complexity was lower than the previous published algorithms. In the same year, Hartenecka et al. [**33**] used the same techniques, but without cross filters between bands that are usually used to remove the aliasing between the bands, instead they monitored the "inband" aliasing in each channel l using a new criterion, $P_i(z)P_i(zW_s^l W_s)\approx 0$, where W_s is a modulation factor and $l = 1,\ldots,S_i\text{-}1$, which is interpreted as the analysis filter, $P_i(z)$, that must not overlap with its modulated version. The limitation of these two methods is in that they are tested in noise-free environments, where it was not clear how their resistances against white and colour inputs would be.

In 2000, Liu et al. [**34**] presented a new technique for non-maximally decimated EPS using a Discrete Fourier Transform (DFT) filter bank. The prototype filter is derived by implementing interpolation of a two-band quadrature mirror filter (QMF) filter. The proposed design was effective in real-time application of the acoustic echo cancellation system but still limited to only high number of bands where the complexity is relatively high. The quadratic constrained prototype [**35**] is used in our proposed scheme instead of the QMF that is used in the latter paper where the perfect reconstruction filter banks are with higher stopband attenuation.

In 2005, affine projection algorithm [**36**] was applied in a multiband adaptive filter scheme with maximal decimation (Section 3.2.5). The drawback in that its performance degrades in the presence of impulse interference [**37**].

In 2007, a new method for an UEPS filter bank design was presented [**38**]. The proposed filter bank consists of different EP sections which are connected together by transition filters. The analysis and synthesis filters are designed separately with each on its own constraints. The results show that the designed filter bank offers a nearly perfect reconstruction performance. The limitation of this work is in its large overall linear magnitude distortion.

In 2008 a new structure [**39**] for decomposition and denoising was proposed. It uses an UEPS filter bank that is constructed from an EP filter bank. The advantage of this method is that it can identify and eliminate the narrow-band noise from the corrupted

signal. The aliasing distortion between adjacent bands has been removed but the amplitude distortion still exists. In our proposed method, we have used octave filter banks that are able to manipulate the low-pass features of the audio signals of the system identification. Chapter 5 presents more details about this issue.

In 2010, Lalitha [**40**] suggested to increase the analysis filter bandwidth to minimise the interband overlapping produced through the downsampling stage. The experiments do not show the real performance of the structure for white noise input in comparison to other proposed methods.

In 2012, Choi et al. [**41**] proposed a robust normalised multiband adaptive filter that is able to detect a noise in the system. The authors produced a new restricted measure in order to decrease Euclidean norms of the variations among the weight vectors. The results show robust behaviour against noise and lower stationary-state misalignment than the standard multiband adaptive filter. The disadvantage of this method is the severe decrease in the convergence performance as the noise increases. On the other hand our proposed algorithm shows better robustness for quick changes of the system identification (see Section 5.7.3).

1.3.4.2 ***Relation of the contribution to the elements of the problem***

a. An adaptation algorithm of an LMS type through normalised step-size is derived. The normalisation, in addition to the use of multiple bands that let our structure work at a minimum sampling rate, a considerable enhancement is achieved in the convergence speed when coloured signals are applied at the input.
b. To reduce the computational complexity, we use:
    a. multiple bands in the separation system that operate at the lowest sampling rates, and
    b. filter banks that are implemented using octave-band with tree-structure and fast Fourier transform.

1.3.4.3 ***Validation of the proposed work:***

To validate our work, we compare our results with the best and most contemporary results [**42**] and [**43**] found in our extensive literature review.

1.3.5 *Second contribution*

A new BSS method is suggested using multiple bands and designed to work in reverberant rooms. The mixed reverberant signal is represented as a convolutive mixture filter bank in a non-maximally decimated (oversampled) UEPS. The filters of the mixture are with real coefficients. The separation system of the proposed structure consists of low order FIR filters. A novel gradient approach of normalisation, of a second order statistics, is used with an adaptation algorithm that is governed by the coefficients of each band of the separation filters. The suggested gradient normalisation provides improved convergence. As the separation filters are constructed with multiple bands at low sampling rates, the proposed approach is less complex than previous methods. Chapter 7 presents a mathematical and theoretical detailed explanation of this contribution.

1.3.5.1 ***Relation to existing methods:***

In 1995, Nguyen-Thi and Christian [**44**] proposed two algorithms for the separation of wide-band signals. These algorithms were made to deal with convolutive mixtures and designed using Finite Impulse Response (FIR) filters. These algorithms are based on fourth-order output cross-cumulants to choose a 2x2 separating system. The two algorithms are proposed using two similar equations that differ only in the location of one element, i.e., Cum31(sr(k), sc(k-i)), and Cum13(sr(k), sc(k-i)). They claimed that the signals are completely recovered for 2x2 mixing system. Sergio and Luis [**45**] studied the stability of the two algorithms assuming temporally independent and identically distributed sources. They showed that, in contrast to the first algorithm, the second one is stable. This proved that the sign and the magnitude of the step size are properly selected. In 1996, Serviere [**46**] used the first algorithm of Nguyen-Thi and Christian to generalise

the source separation problem to convolutive mixtures of wide-band sources. Serviere proposed cancellation of two non-symmetrical cross-cumulants that led to a solution in function of the sources; a good convergence is achieved. The instability, which is studied by Sergio and Luis, of this method is still high; this technique is avoided in our work as we are aiming to improve the convergence rate and keep the complexity low with a stable system. Full details of our method are illustrated in Chapter 7.

The previous works were done for two sources and two outputs. In 1997, [**47**] a multiple inputs multiple outputs linear FIR blind identification system was used with a lower number of inputs than outputs. The experiments were made in a quite (noise-free) environment. The proposed work was based on two Lemmas; the first one is the multiplication of the coefficients of the analysis filters that are multiplied by the corresponding analysis filters that are equal to an identity matrix. The second Lemma connects the mixing matrix to the covariance matrix by a multiplication operation to obtain a diagonal matrix. These two Lemmas were further manipulated to lead to an objective function that is then used in the updated function. The second order statistics is used but higher order statistics was required to completely identify the system. This method showed an almost perfect recovery of the system parameters, but in the cost of high complexity. In our second proposed method, we also used second order statistics in addition to a novel normalisation scheme.

In 2000, Mansour et al. [**48**] used second-order statistics only, to reduce the convolutive mixture to an instantaneous mixture, but the number of sensors were greater than the number of sources. The algorithm proposed some restrictions; among them excluding mixing signals that at a predefined constant rank are irreducible, selecting some column components whose number is restricted to a number of bands with a nonpolynomial matrix. All these limitations made the algorithm tied to these restrictions that are not necessarily valid for other applications. This algorithm with its restrictions does not seem practical to be generalised for BSS applications.

In 2001, Araki et al. [**49**] explained that, the impulse response frequently used in the previous proposed methods was still too long. Their rationale was that the long impulse response restricts the performance of the separation. They suggested minimising

the impulse response length to the length of the FFT frame size at an 8 kHz sampling frequency. They summarised that; the longer the reverberation time, the more difficult it is to reach a good separation performance. Two years later, in 2003, Araki et al. [**50**] changed their approach in respect of the restrictions on the lengths of the room reverberation impulse response and the FFT frame size; they focused on changing the number of bands. They took into account the frequency features of the room and the speech signals. Frequency-dependent multiband-processing is realized in their proposed methods. The only limitation of this work is the long demixing filters that reduce the convergence rate which we have avoided it in our second proposed algorithm. However, we used the same sampling rate and 10 second duration for the speech signals in our experiments (Section 7.2.2).

Kostas et al. published in 2006 [**16**] an interesting paper that proposed a novel idea for BSS in convolutive mixtures of speech based on two methods; a linear prediction method that is used to implement a temporal pre-whitening, and a spatio-temporal separation method that works on entropy maximisation, using a natural gradient algorithm. The linear prediction thoroughly keeps the original spectral features of each source contribution. Combining these two methods, the linear prediction and the natural gradient algorithm in multichannel signal separation framework, led to a novel idea for improving the speed of convergence and the separation performance over existing methods. Furthermore, the linear prediction algorithm has some limitations in terms of the simulation time and memory, when the number of coefficients is large. In our proposed method, we use a natural gradient algorithm with a novel normalisation method.

In 2007, Ghennioui et al. [**51**] proposed an algorithm that is based on the algebraic optimisation of a least mean-square. This algorithm deals with the nonunitary diagonalisation of complex matrices that are usually produced from the blind separation of convolutive mixtures. The algorithm does not need a pre-whitening stage. A new performance index is presented to measure the performance of the separation. The experiments in this work did not show whether there was permutation between the bands. In our proposed method, the experiments show how our algorithm is robust for both the source whitening and permutation problems.

In 2008, Mei et al. [**52**] studied the blind separation for convolutive mixtures. They offered an alternative generalised formulation based on joint diagonalisation (JD). This proposed approach permits us, through a matrix of sources with frequency and time dependent features, to determine some restrictions that when applied, the separation is achievable. This method uses frequency-domain objective functions to measure the degree of diagonalisation of the PSD matrices of the outputs, but the optimised coefficients are the separation-channel parameters in the time domain. The permutation problem is overcome successfully. In comparison to Buchner et al.'s approach [**53**], discussed below, the latter work still offers a better convergence rate and complexity.

Buchner et al. [**53**] used second-order statistics to develop generalised formulations for BSS with convolutive mixtures. Many subsequent researches, including our research, are based on this work. Their approach avoided the internal permutation problem. They optimised the time-lags of the correlations by standardizing the cost function to derive frequency and time domain approaches. Some special cases of these algorithms were studied. A link between the time-domain and frequency-domain versions was presented. The algorithm was able to track the time-varying real acoustic environments by their proposed generalised weighting-function. Our proposal is based on the time-domain part of this approach and improved using multiband processing, octave-band filter banks, tree-structured filter band, and polyphase techniques. The rationales for using the above-mentioned elements and structures in our proposed work are discussed in the coming chapters.

In 2009, Zhu et al. [**54**] combined the natural gradient and temporal complexity approaches to solve the BSS problem of convolutive and temporally correlated mixtures. They used a measure of temporal complexity to retrieve the low frequency components of the source signals. The results proved that their algorithm can separate convolutive mixtures while keeping their structures both in time and frequency domains. The solutions for the scaling and permutation issues are not mentioned in the experimental results and it was unclear how the proposed algorithm will overcome with these two problems that usually arise in the frequency domain. In our approach, we choose the time domain to avoid the scaling and permutation issues.

Ghennioui et al. improved their first algorithm [**51**] by fixing the adaptation step and using spatial quadratic time-frequency spectra [**55**]. The main improvement of this algorithm is its ability to reduce the spatial whitening of the observations with regard to the best reachable performances in the blind sources separation context.

In 2011, Jan et al. [**56**] proposed a new algorithm that consists of the ideal binary mask (IBM) and an independent component analysis using the (ICA), in addition to cepstral domain filtering. Source signals are separated from two mixtures (microphones) using an ICA approach. Then, the IBM is extracted by observing the separated sources where the energy of corresponding time–frequency (T–F) units is calculated in the convolutive ICA algorithm. Finally, the noise produced by (T–F) masking is eliminated using cepstral smoothing. The results show improved speech quality and significantly higher efficiency when it is compared with a latest algorithm. The disadvantage of this algorithm is its limitations to deal with highly reverberant mixtures.

In 2012, Saito et al. [**57**] suggested a novel backward model for BSS by estimating the demixing matrix in frequency domain by solving block by block least squares that approximate to a joint diagonalisation problem. This method does not require solving the scaling ambiguity by other methods due to the scale constraint. The limitation of such frequency domain methods are still in the permutation problem that affects their performances.

In 2013, Fu and Mu [**58**] also suggested an approach that uses the local sparsity speech sources in a transformed domain. This method estimates the frequency mixing system that requires less computation time. The drawback of this method is its weak resistance to the noise signals.

In 2014, Minhas and Gaydecki [**59**] presented a hybrid algorithm for blind source separation that uses a multiple conditions approach to solve the convolutive mixture issue, rather than using a single condition solely. This algorithm employs post-separation speech harmonic alignment that results in a significant enhancement in the performance. The results show that the algorithm requires fewer computations than the best performing methods. The results do not show the performance of this algorithm in a stationary state.

Having introduced the existing methods related to our work, and studying their various advantages and disadvantages; we introduce a BSS algorithm that improves the convergence rate and reduces computation complexity.

1.3.5.2 ***Relation of the contribution to the elements of the problem***

a. To improve the convergence rate, the PSD of the speech signal should be reduced at low frequencies, where the large amount of the speech energy exists. To do that, we break down the input signal into multiple bands by using narrow octave bands [**1**].
b. a novel normalisation is done by reducing the computational complexity of adaptation algorithms

1.3.5.3 ***Validation of the proposed work:***

The fair comparison that we can make, as a validation of our approach, is comparing our results with currently best performing method[1], the Buchner results [**53**], where the time-domain part that we have adopted and improved in our work.

1.3.6 *The relationship between the first and second contributions*

In both methods, a decimated UEPS is used to improve the adaptation convergence. The first method uses an MDUEPS whereas the second uses an NMD (oversampled) UEPS. Both proposed structures offer novel normalisation approaches and octave-band tree-structured filter banks that help in reducing the computational complexity.. The first method was initially designed to solve the BSS problem, but it did not give the expected results. The reason is the aliasing present among non-adjacent bands. However, the first structure, with maximal decimation, was implemented successfully on system identification , as explained in Chapter 5. The first structure is

[1] Our research started in 2012.

modified to work with separating speech signal problems, and resulted in a new (second) structure. The modification done to the first method is the use of UEPS oversampled filter banks that produces a new UEPS structure that is implemented successfully on a BSS. This structure is oversampled and modified from the work in [**53**] , where the signals in every band are downsampled by a factor 1/2 to decrease the effect of overlapping spectrum through the process of adapting the coefficients. The signals are downsampled by a factor of 2 at each band to retrieve the sampling rate of the structure before the last step that recovers the output signal. Different decimation factors are tested for *L*-channel octave-band filter banks.

1.4 Evaluation Method and Experimental Design

The following evaluation methods are used in this work:

a. The ratio between the major and minor eigenvalues is used to evaluate the convergence rate of the multiband adaptive schemes. This evaluation allows us to predict the behaviour of the proposed approach for a given analysis bank. It also reflects the behaviour of the proposed approach to the coloured (highly correlated) signals (Section 5.6).
b. The mean square error evaluation is used (Section 5.6)) to:
   - Compare between different bands of the adaptive scheme.
   - Compare the theoretical and practical estimates of the mean square errors of the proposed multiband adaptive scheme with prototype filters of different orders.
c. The simulation time is used (Section 7.2.2.1) to compare a new proposed scheme with an old one.
d. For performance evaluation, the signal-to-interference ratio is used to evaluate the quality of the full-band and multiband schemes with mixture filters of different sizes (Section 7.2.2.2).

e. Perceptual Evaluation of Speech Quality (PESQ) is used in Section 7.2.2.2 to compare the objective results with some standards and best performing results.
f. The power spectrums are used to prove the robustness of the proposed approach for the whitening sources problem.
g. Computational cost that represents the number of mathematical operations per iteration is used to calculate the number of multiplications per block (Section 6.2.1.2). This measure is used to show the difference between EPS and UEPS.

1.5 List of Publications

[1] S. Badran, S. Ahmadi, D. Menzies and I. Shahin, "Improvement on the BSS for Convolutive Mixtures," in *International Conference on Applied Information and Communications Technology (ICAICT-14)*, Muscat, 2014.

[2] S. Badran, S. Ahmadi, D. Menzies and I. Shahin, "Efficient Separation for Convolutive Mixtures," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, no. 5, pp. 718 - 722, 2014.

[3] S. Badran, S. Ahmadi, P. Bagheri zadeh and I. Shahin, "A Novel Tap Selection Design for Filters in Unequal-Passbands Scheme," in *The Eleventh International Conference on Digital Telecommunications (ICDT 2016)*, Lisbon, 2015.

1.6 Thesis Structure

The thesis is structured as follows:

- Chapter 2 presents a literature review on blind source separation. A brief history is introduced that explains the importance of blind source separation in various disciplines. Blind source separation is classified into several degrees of difficulty depending on the type of the mixing system. A general acoustical description of speech signals is described. A description of the human auditory system and decoding the auditory scene are presented.
- Chapter 3 reviews adaptive filtering algorithms, such as the gradient method, the least-square algorithm (LMS), the excess mean square error (EMSE), and the normalised mean square error (NLMS). The objective function and the recursive equation that is based on the gradient descent method are derived. The updating equations of the adaptive filter coefficients of the different types are extracted.
- Chapter 4 presents a review of the important topics of the filter bank and multirate systems used for the development of this work. In this chapter, an MD_UEPS with random bands and a non-maximally decimated filter bank is discussed, as they are required to explain the proposed method in Chapter 5. A maximally decimated tree-structured filter bank and its equivalent representation are introduced. Polyphase representation is applied and studied for both EP and UEPS filter Banks, and then cosine-modulated filter bank is presented.
- In Chapter 5, an MDUEPS with arbitrary number of bands is proposed. The first contribution of this research work is the derivation of the MD_UEPS from UEPS without decimation, which employs an analysis bank and filters with non-zero coefficients that are used to construct an equivalent FIR system. The convergence, optimal coefficients, mean square error, and computational cost are investigated.
- Chapter 6 presents the BSS problem of determined instantaneous mixtures . The ICA is again mentioned to clarify the stochastic gradient of the objective

function (entropy) and the natural gradient. BSS for convolutive mixtures is presented. Frequency and time domain BSS methods are studied. The permutation and scaling problems are highlighted again. Various performance measures are demonstrated.

- Chapter 7 proposes adaptive multiband schemes for recovery of the sources, using UEPS subsampled filter banks. This approach uses separation filters with various sizes and is adapted at different rates. The adaptation is achieved using algorithms based on a natural gradient. Experiments are presented that show reduction in computational cost and increase in the SIR compared to the full-band scheme.
- Chapter 8 concludes this work, discusses its limitations, and suggests some ideas for future work.

Chapter 2
Literature Review

This chapter presents a literature review of the fundamental ideas of audio processing in blind source separation (BSS), and the general modern techniques and applications of the BSS. The basic characteristics of speech sounds are discussed in detail. Moreover, the principles of acoustics are also explored. Tables and plots are presented in order to explain the design of a specific room, taking into consideration many parameters, such as the noise sources, the frequency limitations in different situations, and the materials used in the room. The relationship between the reverberation time and the room volume is also presented. Some metrics that are usually used in sounds are discussed; among them is the speech intelligibility level against the distance from the speaker to listener. The room quality and the description of each quality level are described. The impulse response techniques and how it is constructed are discussed in depth. All this and more details are presented to give us a full understanding of the kind of rooms that are usually used in the research world. Standard virtual room scenery that is used in our work is presented. Using this standard room will allow us to validate our results with a corresponding work that uses the same scenery. Then, state-of-the-art in the BSS work is presented. Applications for different BSS methods are discussed. The basic principle of BSS is explained. The time- and frequency-domain BSS methods are discussed. The main metrics that are used to test the quality of the speech are presented.

Chapter 3 and 4 are a continuation of our literature review. They are separated from Chapter 2 to highlight the literature review on the most important parts of our proposals. Chapter 3 describes the fundamental discussion of the adaptive algorithms, where our new adaptation method is developed with a novel normalisation that improves the convergence speed. Chapter 4 presents the most important structures that are used and developed in our structures, and/or used for comparison with our proposed structures.

Usually, statistics and computational approaches are employed to estimate the original sources, using ICA. This method is capable of extracting the original signals from the mixed signals using the non-Gaussian and linear mixture modelling.

The problems concerning blind source separation vis-à-vis linear instantaneous determined mixtures (hereinafter referred to as instantaneous mixture) have long been resolved. However, a major challenge is to find a good solution for the problem of linear convolutive determined mixtures (referred to as convolutive mixtures) which involves delays and reverberation. An extensive research has been done in this area and there are essentially three ways to treat the problem of convolutive mixtures: in the frequency domain [**60**], in the time domain [**61**], and the hybrid solution [**62**]. The problem of the time domain solutions is the reverberation time which is the reason behind the high computations. The order of the reverberation time is about 0.25 seconds for a medium size room that is approximately 3.5 meters in width by 4.5 meters in depth by 2.5 meters in height. For samples sampled at a frequency of 8 kHz, for example, we would have to model separation filters in the order of 2000 coefficients (Assuming the separation filters are in the same order as the mixture filters.). Converting the convolution in the time domain into the frequency domain will help reduce computations as this conversion allows working with multiplications rather than convolution. Thus, the problem can be seen as a case of an instantaneous mixture in each frequency component. However, it should be borne in mind that in practice, there are two disadvantages of this conversion. The first one is that such transformation is only possible in the discrete domain when the convolution is circular, not linear. The second disadvantage is the need for the use of large blocks of signal samples, which consequently necessitate the use of long filters. In this case, due to the non-stationarity of the acoustic environment (mixed system) and audio signals, there are not enough samples in the frequency domain for a proper estimation of statistics in each frequency bin. These drawbacks can severely limit the performance of these algorithms. Some solutions were presented by the possibility of working with independent separation filters and a reduced order on each band. These systems typically use banks with complex coefficients [**63**], [**64**].

2.1 Characteristics of speech sounds

This section discusses the basic acoustic characteristics of speech signals and explains them in terms of spectrograms and spectra [**65**].

2.1.1 *Vowels*

Vowels are generated with open vocal tract as the air is not totally blocked. Thus, the voice is somewhat loud. Vowels can also be generated with a vocal fold vibration. The main acoustic characteristic of vowels is the position of the formant frequencies (F1, F2, and F3). The formation of the vocal tract defines the position of the formant frequencies. Variations in the location of the articulators will improve the formation of the vocal tract and consequently the position of the formant frequencies. As the same formant frequencies can be produced with multiple articulatory positions; the formant frequency position is a critical limit on vowel quality, instead of the locations of the articulators. For a speaker or a set of speakers that have similar vocal tract length, every vowel is inherited with an individual acoustic formant frequency pattern. The spectrograms of a male English speaker who uttered the vowels [a] and [i] are exhibited in Fig. 2.1 and Fig. 2.2 [**66**].

2.1.2 *Consonants*

Consonants include a changing of the airstream, from a slight change in the approximants case, to a sharp change in the plosives case. In this section, we present the acoustic characteristics of consonants, starting with the closest one to the vowels, that is, approximants, and ending with plosives [**67**].

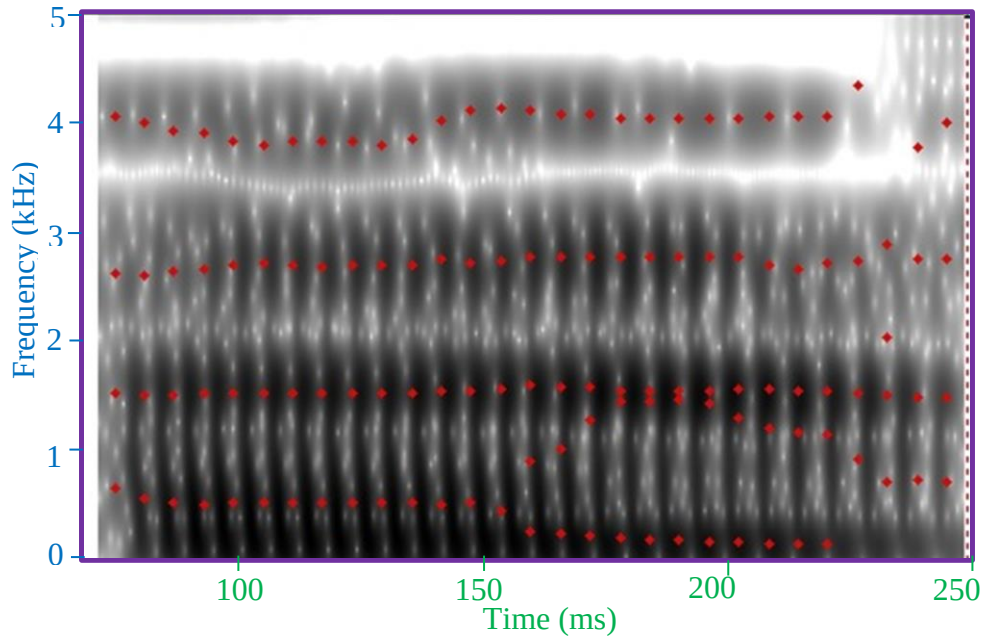


Fig. 2.1 Spectrogram of [a], uttered by a male English speaker

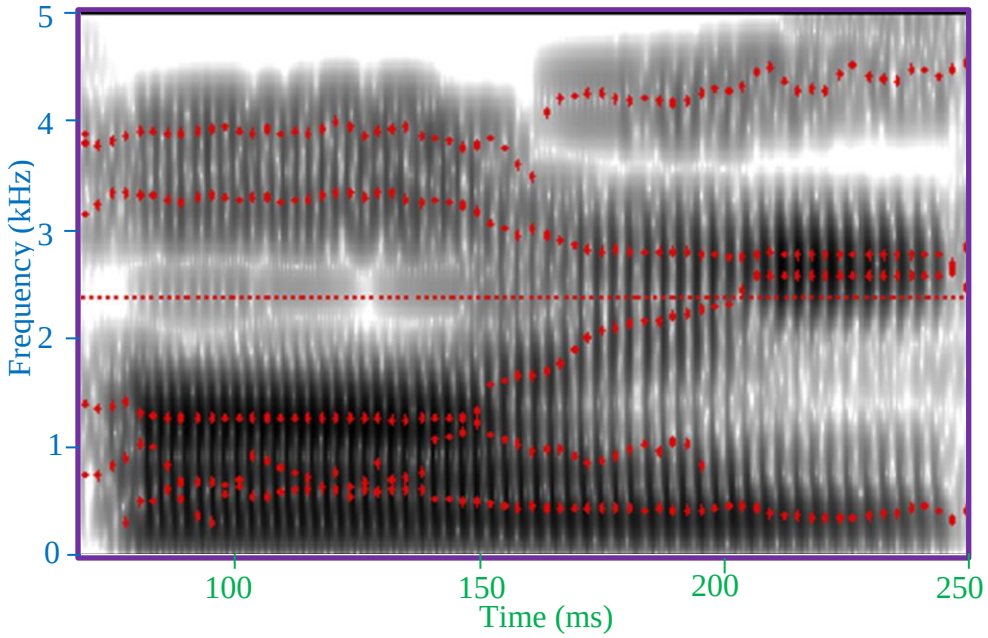


Fig. 2.2 Spectrogram of [i], uttered by a male English speaker

2.1.2.1 *(Central) Approximants*

Approximants happen when two articulators get close to each other without preventing the flow of air. Therefore, the acoustic features are very similar to the vowel features generated at a similar position in the vocal tract. Their formant shape is clear but slightly lower than that for the vowels as the approximants' somewhat larger constriction leads to a smaller steady-state part and poorer acoustic energy. For instance, the approximant [j] is very similar to the vowel [i], but the blade of the tongue is nearer to palatal approximant for [j].

2.1.2.2 *Fricatives*

Fricatives are created by a thin contraction in the mouth. The turmoil causing by air crossing over this contraction is the typical sound source for all fricatives. Fricatives can be described in four features: spectral properties of the transition into and out of the adjacent vowels, spectral properties of the friction noise, duration of the noise, and amplitude of the noise. The total spectral shape of every fricative is typically specified by the shape and size of the oral cavity in front of the contraction. The filter modelling of speech construction, explained through the acoustic characteristics of vowels, also describes the acoustic features of consonants [**68**]. As obstruents are formed with a strong contraction, their resonant frequencies can be computed for the front and back cavities individually, without any complication.

2.1.2.3 *Plosives*

Plosives cause the air flow to be interrupted by a brief contraction of the oral cavity. The freeing of the contraction and the next movement of the articulators to the subsequent sound leads to a burst frequency and formant transitions. Acoustic characteristics of intervocalic plosives are shown in the spectrograms of Fig. 2.3. The presence of a gap in the spectrogram is due to the fact that the locked part of the

consonant does not have energy in the cases of voice bar (very low-frequency energy) in wholly voiced plosives and the voiceless plosives [**65**].

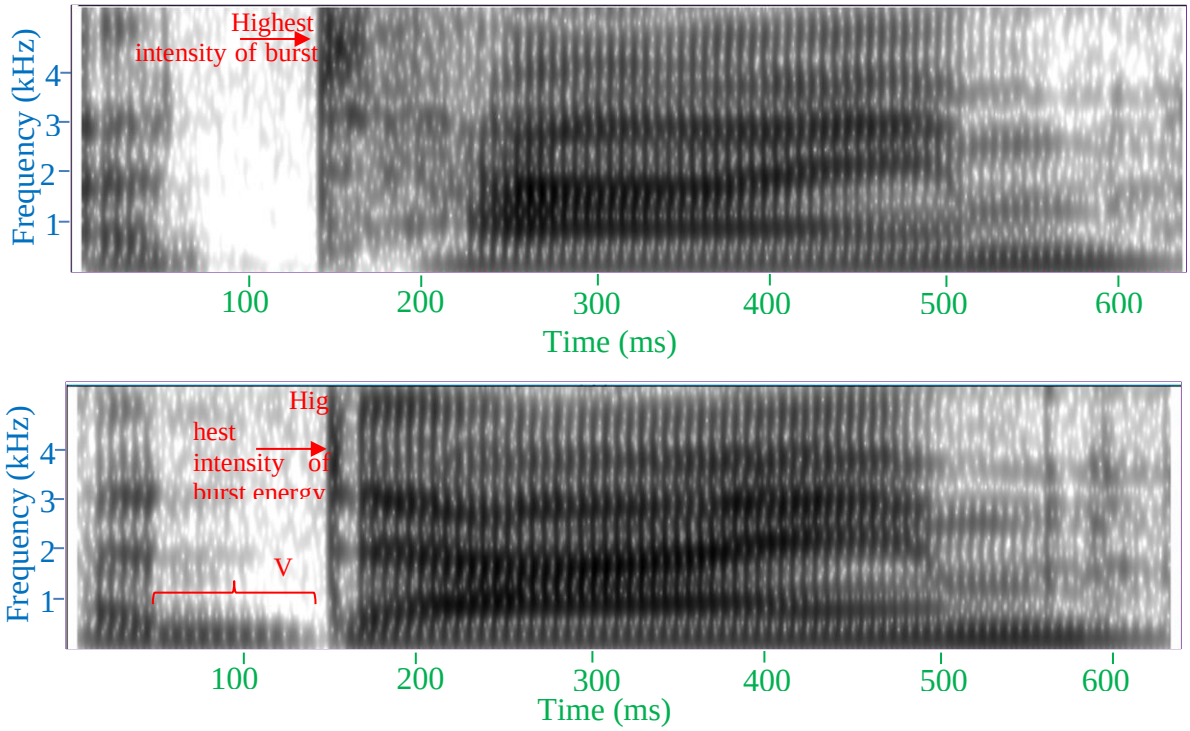


Fig. 2.3 Spectrograms of a female English speaker (a) *a time* with a voiceless and (b) *a dime* with a voiced intervocalic plosive.

2.1.2.4 ***Nasals***

In the nasal consonants case, the airstream is momentarily impeded by the contraction in the oral cavity. On the other hand, the velum is depressed in a way that air can get out only via the nasal cavity. The produced sound is called a nasal murmur. The nasal and oral resonance cavities need to be joined in a complicated way to produce Nasals. Nasals are characterised by low F1 formant amplitude that is also known as the nasal formant. This reason is because the air stream is blocked by a thin hole in the nasal cavity.

2.1.2.5 *Lateral approximants*

The lateral alveolar approximant [l] and the nasals characteristics are almost the same. During the generation of [l], a portion of the tongue touches the top of the mouth. The formant frequencies of a male for [l] are roughly as follows: F1 = 0.34 kHz, F2 = 1.2 kHz and F3 = 2.8 kHz. Fig. 2.4 illustrates the characteristics of the lateral approximants. A small pocket of air is present over the tongue while air is expelled from both sides of the contraction [**65**]. It is the same scenario for the nasals where the vocal tract can be seen as having a leading tube and a side tube. The leading tube lengthens glottis to mouth opening, whereas the pocket of air is exhibited as a short side tube. The joint between the leading and side tubes leads to anti-formants. Formants from the side tube will wipe out formants from the leading tube. As the side tube is small, the anti-formant will be large, 2 to 2.3 kHz for a male.

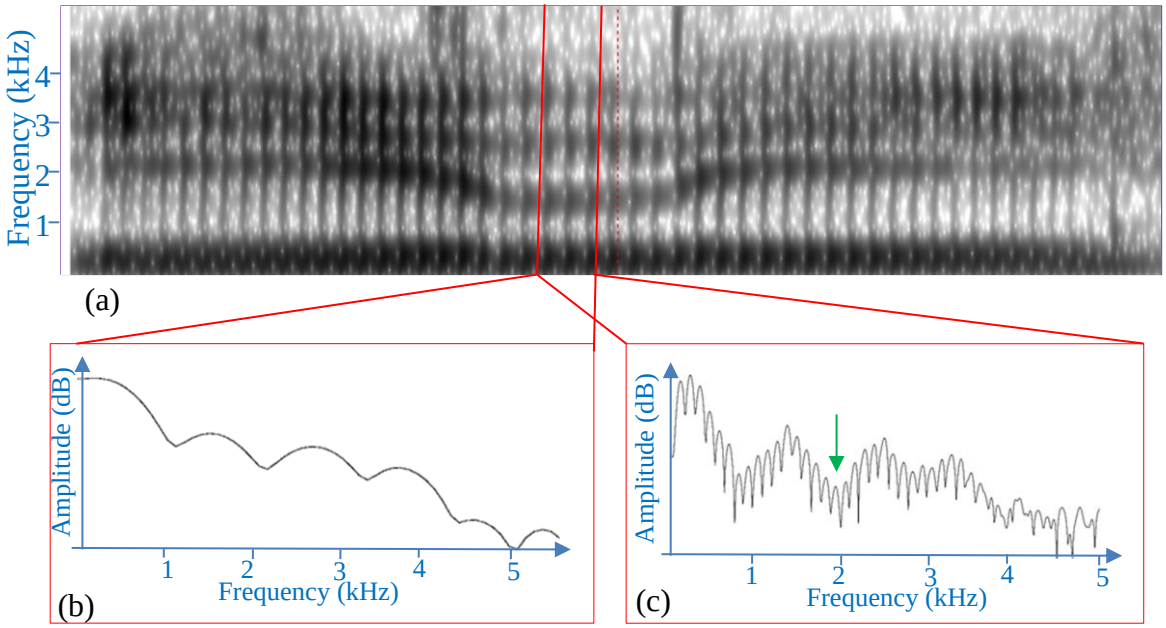


Fig. 2.4 (a) Spectrogram of the utterance [ili] together with (b) LPC and (c) FFT spectra of the central part of [I], spoken by a male native speaker of English. The arrow in the FFT spectrum points to the anti-formant at 2000 Hz.

2.1.2.6 *Affricates*

Affricates have similar characteristics to both plosive and fricative, respectively. Affricates differ from fricatives in the long rise time, as the temporal interval starts from the beginning of the consonant to the maximum amplitude of the friction noise. However, affricates have a short rise time [**68**].

2.2 Acoustics: Concepts and Perception

This section provides an in-depth discussion of room acoustics, general room-acoustical characteristics such as reverberation time, the room impulse response, and different source-receiver configurations. The source of noise can be both inside and outside of a building. Therefore, some acoustic factors should be analysed, such as the location of the source and the receiver. A closed environment causes some acoustic phenomena, such as reverberation as the sound waves encounter physical barriers of propagation. However, this is not the case in free environments.

2.2.1.1 *Reverberation Time (RT)*

Reverberation time refers to the time taken for the sound to decay by 60 dB, after the sound source has ceased. This is certainly one of the most important criterion for analysing a room. Sound does not die the moment it is produced; instead, it continues to be heard for a few moments due to its reflection [**69**].

Physically, an acoustic room is characterised by its impulsive response, which is calculated by the source-receiver pair [**70**]. In other words, the response that reaches a given receiver positioned in the environment, which in turn suffers a source disturbance, completely describes the acoustic system between these two points and therefore serves as a good basis for interpretation of the acoustic room. . A typical example of Room Impulse Response or Energy Decay Curve is shown in Fig. 2.5 [**71**].

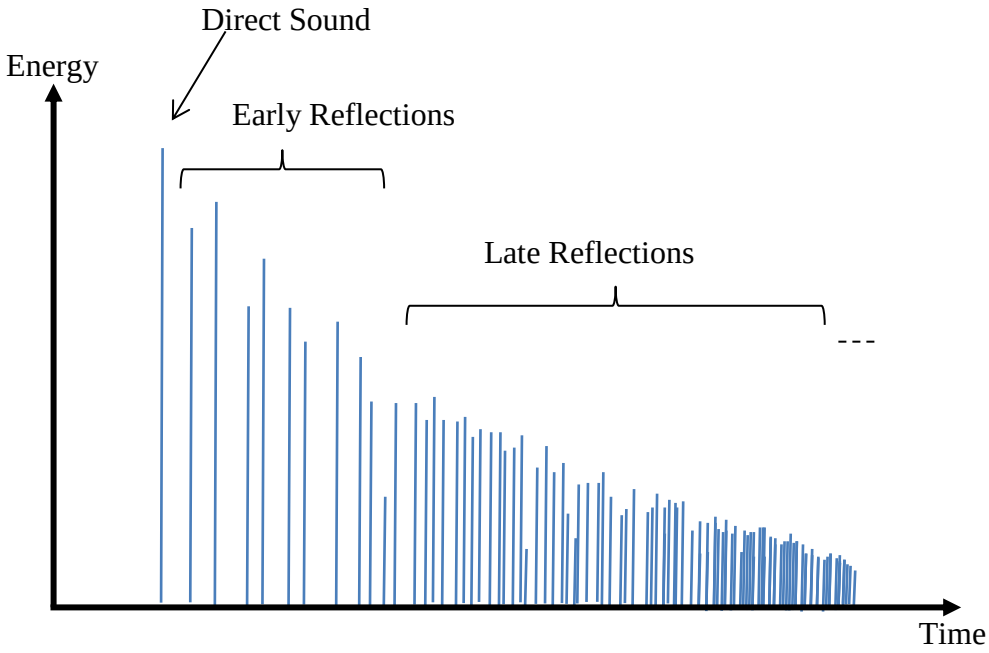


Fig. 2.5: Decay energy curve with respect to time, Room Impulsive Response

When analysing Fig. 2.5, it can be observed that the components in the energy decay curve are arranged to characterise a series of pulses. It is still possible to highlight three distinct phases: direct sound, early reflections and the reverberant tail. The first impulse, direct sound, is characterised by the signal emitted by the source and that reaches the receiver without suffering distortion from the environment. Early reflections help to determine the perception of the direction of the sound incidence and the reflections.

2.2.1.2 *Speech intelligibility test*

The term "intelligibility" refers to the quality of perception of some language units (phonemes, words, sentences, etc., depending on the method). The percentage of units correctly understood in the respective spoken units is called the "intelligibility index [**72**].

Speech intelligibility is a parameter used to evaluate the correct speech perception by the listener. The speaker-listener tests are a direct means of obtaining the speech intelligibility in the room [**73**].

The test consists of pronouncing a series of monosyllables representative of the language. This pronunciation can be done through an elaborate recording in an appropriate place or by a present speaker. The reading rhythm should be a monosyllable word every four seconds. The result of correct words is the average percentage of the monosyllables understood [**74**]. Speakers should be homogeneously distributed in the room, and each speaker should be present in only one room to prevent words from being memorised. The height of the speaker in relation to the listeners, the distance between the speaker and the listener, the speed of the reading, the possibility of lip reading are irrelevant to the results [**74**].

The words should be monosyllables without meanings, but phonetically balanced and representative in the specific language, and the monosyllables should be composed of consonant-vowel-consonant [**75**].

The mean of the intelligibility test results shows the percentage that indicates the intelligibility of the environment. If the speech intelligibility in a space is less than 90%, acoustic processing mustbe performed to minimize the reverberation and enhance the SNR [**73**], see equation (2.5). Table 2.1 shows the scale and description of the quality of the intelligibility in the room [**72**].

Table 2.1: Scale and description of intelligibility quality

Room Quality	**Quality description**
Excellent	Very good perception without any effort to listen
Good	Good perception; Easy understanding of words
Acceptable	Understanding is difficult; Some effort is needed to listen
Poor	The listener has difficulty understanding and reconstructing words and sentences
Bad	It is impossible to understand, therefore, to reconstruct words and sentences (simply regular)

2.2.1.2.1 Virtual Scenario

A virtual scenario (proposed competition for separation of signals published in 2012 in [**76**] and shown in Fig. 2.6. The sound sources are placed 1 m away from the midpoint between the sensors in two different directions: -500 and 450. This scenario was proposed for a competition promoted in 2007 by the SiSEC-2006[2], with the results presented [**77**] in a panel of discussions during the 7th International Conference on Independent Component Analysis and Signal Separation (ICA 2007); the results are updated in [**76**]. The scenario had N = 4 distinct sources in different directions. This scenario will be adopted in our experiments in Chapter 7 (Section 7.2.2); however, in our experiments to reduce the complexity of the BSS procedure, we will reduce it to only two sources.

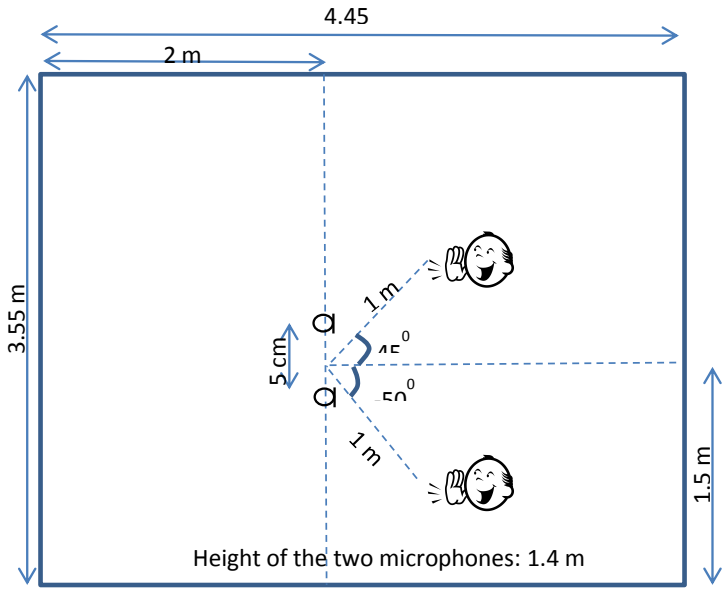


Fig. 2.6: A Virtual Room Scenario.

[2] Signal Separation Evaluation Campaign Community

2.3 Research and State-of-the-Art in the Source Separation Area

The speech signal can be partially or completely hidden by noise or other speech signals. A major operation in the voice communication systems is the extraction of the original speech signal from the corrupted signal by additive noise or other speech signals.

In many situations, the signal received on a set of sensors is the sum of many contributions called "signal sources". In general, it is realistic to estimate which sources provide independent stochastic signals. To study the effects of each source on all signals, it is necessary to separate the sources by a filtering procedure. To achieve this separation, a partial identification of sources (usually not total) must be made.

Significant work has also been done on the source separation problem using the statistical properties of the second order signals. Several methods using second order statistics of the signal have been developed to identify and detect the number of sources.

Cardoso and Souloumiac [**78**] has proposed joint approximate diagonalization of eigenmatrices (JADE) method. JADE is an analysis of algebraic methods based on cumulants showing how high order correlations can be efficiently explored to find independent components. This algorithm is called Jacobi's Algorithm due to the fact that we try to maximize the measures of independence through the Jacobi diagonalization method. The independence measures used are based on fourth-order cumulants. The advantage in relation to other algorithms is that one can modify the training parameters in large steps without having problems of convergence. The great advantages of this method when compared to gradient-based methods are, that there are no parameters to be tuned in the basic implementation and complex learning algorithms are not required. Comon [**15**] a few years later demonstrated the validity of these methods. This work then inspired Belouchrani et al to propose the Second Order Blind Identification (SOBI) algorithm [**79**], in which it is exploited not only one but several covariance matrices of observations associated with non-zero delays. Rather, they show that, after observations of whitening,

a joint diagonalization of matrices in question can estimate the mixture to a near-trivial matrix. However, this approach requires that the sources have distinct spectra.

In the case of a reverb environment Douglas [**80**] proposed an approach that is called Spatio-Temporal FastICA. This is a temporal method in which the authors combine a pre-whitening spatiotemporal multichannel via several steps of linear prediction of least squares with adaptive procedures.

A BSS based on perceptual filter bank and ICA was proposed by Missaoui and Lachiri [**81**]. This is a blind separation method that employs the kurtosis maximization criterion (KMC) and mixes two approaches: the perceptual filter bank (PFB) and the ICA. The PFB was constructed by regulating an nondecimated decomposition tree to deal with the features of the acoustic model. This method suggested a technique that converts the observations signals into a new representation by employing the KMC. The goal of this conversion is to make the non-Gaussianity higher that is necessary for independent component analysis. The achieved results illustrat that the proposed approach offers a significant enhancement in contrast with with some standard techniques.

The previous three approaches SOBI, JADE, and FastICA became standard benchmarks in the blind source separation. We have used these approaches in addition to Missaoui and Lachiri method [**81**] to validate the results of our second proposed work, see Section 1.3.4.3.

2.4 Principle of the Ideal Separation

The principle of separation can be summarised in Fig. 2.7. It is found in the ideal case as a filter matrix $W(z)$ of size $N \times M$ which inverts the mixture, i.e. which provides the output vector:

$$\tilde{u}(n) = [W(z)]x(k) = [W(z)][A(z)]u(k) \approx u(k) \qquad (2.1)$$

The filter $W(z)$ is:

$$W(z) = A(z)^{-1} \qquad (2.2)$$

The general principle used to determine the filter is based on a criterion of independence. Indeed, if the sources u_j are mutually independent, then any linear combination of a filtered version of two or more sources induces dependent signals between them. Thus, for this ideal separation to be achieved, $W(z)$ must be selected, as the different outputs $y_i(k)$ are mutually independent. This concept is the basis of the independent Component Analysis [**82**] and is widely practiced in the instantaneous mixtures [**83**].

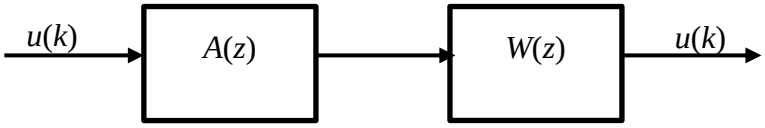


Fig. 2.7 Principle of the separation of convolutive mixed sources.

2.5 Choice between Frequency Domain and Time Domain

In fields such as audio applications, filters are estimated with more than one hundred coefficients. Calculations induced by convolution relationships can then quickly become expensive. However, the convolution relationship becomes a simple product in the frequency domain in which it is easy to see the advantage of the speed of the FFT algorithm.

Thus, after passing through the frequency domain, the matrix $A(f)$ of equation (1.6) is constant for a given f, which provides an instantaneous linear mixing of the contribution of the N sources . By separating the sources in each of these bands, it is then possible to retrieve the original signal by inverse Fourier transform and summing the outputs associated with each band. The ICA conventional methods is able to solve such a problem, i.e. applicable to linear instantaneous mixtures.

Some work has been done using narrow band [**84**]. Other studies, however, deal with the problem with the use of broadband. Jun [**85**] proposed, for example a bi-spectral approach but he did not give details of the reconstruction source once the separation was performed in the frequency domain. Capdevielle et al. ([**86**] and [**87**]) have presented an

approach that utilises the continuity of the spectrum between different frequency bands. This solution often used in some available frequency approaches, presents certain restrictions on the source signals. The Fourier transform tends to gaussian signals, which does not facilitate their separation by conventional methods, except for special cases [**86**] and [**87**]. Thus, the frequency approach, despite its simplicity, has a problem of indeterminacy that could lead to an erroneous reconstruction of sources without realizing it.

Araki et al. [**50**] explained that when the convolutive BSS is performed in the frequency domain, a tradoff must be taken into account at the windowing time between the frequency transformation and the reverse transformation. If the window is too long, there there will be less frequency samples of products and the independence assumption will be invalid between the sequences of spectral samples from different sources. If the window is too short, there is a lack of frequency resolution and temporal filters to the separation matrix will be too short.

Attias and Schreiner [**88**] presented a very interesting approach that can combine the information in two domains, frequency and timeboth the time. However, it is limited to super-Gaussian signals, which is generally not the case for audio and speech signals.

Wu and Principe [**89**] and Parra and Spence [**90**] also showed oriented time-frequency approaches in which the indeterminacy is removed by a constraint on the length of the temporal reconstruction filters in relation to the frequency resolution. This condition eliminates indeterminations when signals are being restricted and in particular non-stationary. This is not always the case, for example in the case of communication signals taken at the symbol frequency. Note, however, that this is a special case and that this type of signals is generally cyclo-stationary.

Because of all of the limitations that are discussed in the aforementioned previous methods, we decided to develop a method based on the separation of the sources in the time domain. Multiband signals are decomposed to decrease the input correlation. This decomposition increases the convergence rate.

2.6 Ambiguity Separation

Separate blind sources contained in mixtures thereof are tainted by ambiguity. Indeed, the separation can be achieved with an abundance of solutions [**91**], [**92**]. For each case, instantaneous or convolutive, the existence of ambiguity affects the reconstruction of sources.

2.6.1 *Ambiguity Permutation*

An estimate of the inverse matrix of filters is not unique. This can be seen on the mixing model (instantaneous case) for example, by swapping two columns in the mixing matrix; then using the commutative property of multiplication operation and the addition operation, we can write:

$$x = \sum_{j=1}^{p} u_j a_j = u_1 a_1 + u_2 a_2 + \Lambda + u_M a_M = u_2 a_2 + u_1 a_1 + \Lambda + u_M a_M \qquad (2.3)$$

We note, based on the initial mixture, and compared with the initial vector representing the sources, there has been an order changing between sources of index 1 and 2, while maintaining the same vector mixtures. This causes a permutation of the first line with the second in the mixing matrix. We can conclude that the permutation does not change mixtures, but implies that the order of sources is unknown.

2.6.2 *Ambiguity Scale*

We saw in the previous subsection that the permutation of two columns in the mixing matrix and two sources does not change the mixtures. We will see in this section, the multiplication of a column (the mixing matrix) and a division by a scalar source will not change the vector mixtures [**91**]:

$$u_1 a_1 + u_2 a_2 + \Lambda + u_M a_M = \left(\frac{1}{\alpha_1} u_1\right)(\alpha_1 a_1) + \left(\frac{1}{\alpha_2} u_2\right)(\alpha_2 a_2) + \Lambda + \left(\frac{1}{\alpha_M} u_M\right)(\alpha_M a_M) \qquad (2.4)$$

2.6.3 *Filtering Ambiguity*

In most cases, the convolutive mixture is reduced to a set of instantaneous separation while passing through the Fourier transform domain. Thus, ambiguities can occur at each frequency that involves filtering signals in time. However, this should not corrupt the process of source separation. The method of Minimal Distortion Principle (MDP) is commonly used to counteract the filter for the frequency ambiguity approaches [**93**]. Accordingly, separation of convolutive mixtures can be accomplished with filtering almost without jeopardizing the separation process.

2.6.4 *Complete Ambiguity*

In the case where the ambiguities are associated, the blind source separation, in addition to estimation errors, cannot recover the source signals in the linear mixtures but for a possible permutation and/or a possible scale factor. In the case of convolutive mixtures, all ambiguities can be modelled by a complex operation and the original sources will be estimated by a signal filtering.

2.7 Evaluation of the Source Separation

The definition of evaluation criteria is an important step in comparing different algorithms for source separation. In this section, we present different criteria for comparison and measurements.

2.7.1 *Criteria for Comparison*

A message is usually a carrier of information, and speech is no exception to this principle. However, this information may be partially or completely obscured by noise. The quality of a message can set a criterion of a good message reception against the message with environmental attacks. However, the quality is generally quite a subjective

notion that it is necessary to recall the methods to quantify. The criteria for quality of a message can be divided into two groups; the objective criteria and subjective criteria.

2.7.1.1 *Subjective Quality Criteria*

Subjective criteria are judgments made by the auditor about the message it receives. These criteria are referred to as subjective because they involve one of the two communication speakers, in this case the receiver, which has no prior no knowledge of the message that is sent. An example is the subjective Mean Opinion Score (MOS) method [**94**].

In a subjective test, participants are asked to test a system under different conditions and to note on a quality scale, the voice quality of this system. In general, the quality depends on the person who judges. His/her perception involves past experience, expectations and their emotional state.

Voice quality is dependent on the person who evaluates. Thus, the notes of the participants for a given test condition are averaged to obtain the mean opinion score, which reduces the subjective effect on the assessment of voice quality. In addition, the perception of voice quality depends on the context and environment in which the person who judges is placed. Indeed, if it is simply listening to a voice message (listening context) or if involved in a conversation with someone (conversation context), attentional processes involved are not the same and judgment of quality is affected. Similarly, the environment (noise, additional visual or audio information, etc.) influences the judgment of quality. Thus, the conditions to be tested are defined in terms of the objective; participants are likely to change and adapt to the environment.

The scoring is conducted according to the methods defined by the International Telecommunication Union (ITU) in Recommendation P.800 [**94**], with the categories:

Bad = 1, Poor = 2, Fair = 3, Good = 4, Excellent = 5, see Section 0.

2.7.1.2 ***Objective Quality Criteria***

Objective criteria are measures that take into account the message as delivered by the transmitter and is understood by the receiver. The objective criteria are accurate, despite the existence of some errors that may appear due to the method of approximations that is applied in calculations. These criteria rely on dissimilarity measures on signal and involve the message at the transmitter and at the receiver. In order to compare the different experimental results as objective methods, we present here four different metrics: time domain, the spectrogram, the SNR and the PESQ (Perceptual Evaluation Signal Quality).

2.7.2 *Time Domain Measurements*

Measurements in the time domain are mainly intended for encoding/transmission systems that attempt to reproduce the original waveform (for example, in the case of waveform coding techniques). The simplest and most common time domain measurements for signal quality evaluation are Signal-to-Noise-Ratio (SNR) and Segmental Signal-to-Noise-Ratio (SSNR).

Let $x(k)$ be the original speech and $y(k)$ be the deformed version of the signal, the SNR is defined by

$$SNR = 10log_{10} \frac{\sum_{k=1}^{N} x^2(k)}{\sum_{k=1}^{N} (x(k) - y(k))^2} \tag{2.5}$$

where k is the time index covering the measurement period.

SSNR is a variation of the SNR which operates on short segments (15 to 20 milliseconds). This facilitates time alignment and provides results that are slightly better (in terms of correlation with subjective evaluation) than those of the SNR. It is calculated as follows:

$$SSNR = \frac{1}{N}\sum_{m=1}^{N} SNR_m \quad (2.6)$$

This is an average of the SNR values obtained for separate frames. The frame is a block of samples.

Measurements in the time domain are easy to implement and can be useful for detecting distortions introduced by additive noise or noise generated by the waveform encoder. However, they show limitations when they are used in a general context, especially when there are degradations, such as filtering or phase distortions.

2.7.3 *The Spectrogram*

The spectrogram is a combination diagram at each time *t* of a signal with its frequency spectrum. It is used to identify sounds and it is widely used in the field of speech recognition.

This is a display tool that uses the technique of Fourier Transform and therefore the calculation of spectra. It began to be widely used in 1947, to the appearance of sonograph, and became an essential tool for phonetic studies for many years.

The spectrogram allows us to highlight the different frequency components of the signal at a given time. A fast Fourier transform is regularly calculated at short time intervals. In this analysis method, the signal is considered as indefinitely stable and consists of an unchanging sum of sine functions of different frequencies. To overcome this theoretical constraint of invariability of the signal, it is necessary to convolve the signal with a suitable time window, since each spectrum computation requires convolving the signal with the time window at a particular time.

Different time windows exist but each introduces a greater or lesser residual error in the spectrum obtained as a result of the chosen form, which may, in the worst case, be triangle or square. The choice of the window size in a number of convolution of points is also important with respect to the quality of the obtained frequency analysis.

After successful convolution, the Fourier transform is computed on the entire window, the rest of the "signal" is then equal to 0. This process allows us to obtain a spectrum which corresponds to a frame, a set of frames calculated at regular intervals to obtain the desired spectrogram.

2.7.4 *The Signal-to-Interference Ratio (SIR)*

This criterion is represents the interference between an undesired signal and the required source signal. This measure was used in evaluating our experimental results and further details on the calculation of this ratio and its types will be explained in Chapter 6.

2.7.5 *Perceptual Evaluation of Speech Quality (PESQ)*

This is a tool that measures the performance of the algorithms in terms of the perceived quality of received speech. It is an objective measurement that evaluates the quality of the speech using PESQ tool. It compares the received signal and the original signal and assigns a score Mean Opinion Score (MOS) between 0 and 4.5. This tool is used in our experiment in Section 7.2.2.2.

2.8 Conclusion

This chapter presents a detailed introduction to the characteristics of sound and acoustics principles. Moreover, the concept of BSS is introduced. We have reviewed the main foundations for the development of approaches to dealing with the problem of blind source separation. In relation to the different methods presented in this chapter, we will focus on the problem of determined convolutive mixture, breaking down the inputs into bands using a maximally decimated filter bank. Through breaking down the signals and decimation of the signals observed, the computational cost will be decreased. One challenge will be to deal with the permutation between signals in different bands to recover correctly the estimated signal sources and without degradation of performance.

Chapter 3
Adaptive Algorithms

3.1 Introduction

This chapter presents a background literature discussion of adaptive algorithms. Adaptive algorithms have become a very valuable tool in signal processing because of their ability to modify the behaviour of a system over time, based on some performance criteria. This chapter reviews adaptive methods based on the gradient descent technique, which will be used to derive the adaptive UEPS presented in Chapter 5. The adaptive UEPS will also be used in the blind source separation algorithms discussed in Chapter 6. Then we present the Least Mean-Square and Normalised Least Mean Square (NLMS) algorithms that are based on stochastic gradient method [**1**], and that will be used for comparison of performance with the adaptive algorithm proposed in Chapter 5. Affine Projection method, The Time-Varying LMS, and recursive least-square method are also discussed in this chapter.

3.2 Algorithms for Adaptive Filtering

The adaptive filter has been extensively studied in recent decades due to its flexibility and applicability, as it self-adjusts the filter coefficients according to an adaptive algorithm.

3.2.1 *System Identification*

In this application, $x(k)$ is considered as a white noise that is applied to the adaptive filter and the unknown system [**95**]. When the mean squared error of the scheme

is minimised, the adaptive filter represents the unknown system model. If the adaptive filter in Fig. 3.1 has *K* coefficients, its input vectors and coefficients can be defined, respectively, as:

$$x(k) = [x(k) \Lambda \; x(k-(K-1))]^T \tag{3.1}$$

$$\omega(k) = [\omega_0(k) \;\; \Lambda \;\; \omega_{K-1}(k)]^T \tag{3.2}$$

where $\omega(k)$ are the coefficients of the adaptive filter. Then,

$$y(k) = \omega^T(k)x(k) \tag{3.3}$$

is the output signal of this filter

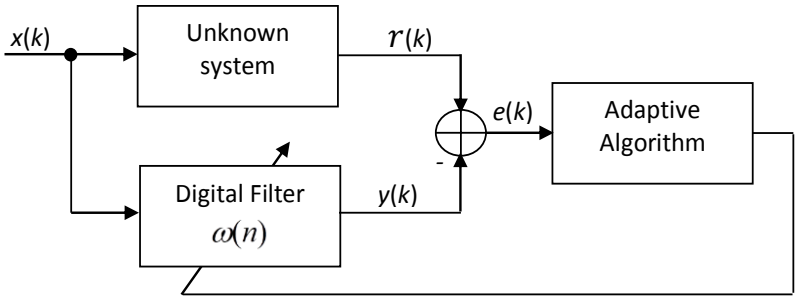


Fig. 3.1 System identification.

The disadvantages of the conventional system identification approaches have motivated the use of multirate techniques for system identification [**1**]. System identification with multirate techniques is computationally efficient and the use of distinct bands improves the convergence rate. The following is a study of some known adaptive algorithms.

3.2.2 *Gradient Method*

The gradient method is a numerical method that is widely used to modify the adaptive filter coefficients $\omega(k)$. Fig. 3.2 illustrates how the objective function can be

calculated. Each iteration is based on minimising a given objective function $\mathfrak{J}(k)$. For supervised adaptive filtering, this function is based on the error between the reference signal $r(k)$ and the output signal $y(k)$ as shown in Fig. 3.1. BSS that is based on the objective function does not use a reference signal, but instead, the mutual information at the output of the system. The procedure for finding the minimum value of objective function $\mathfrak{J}_{\min}$ will be chosen as shown in Fig. 3.2. The recursive update equation of the coefficients is given below:

$$\omega(k+1) = \omega(k) + \gamma[-\nabla(\mathfrak{J}(k))] \tag{3.4}$$

where $\omega(k)$ is the filter coefficients, γ is a real positive constant called the adaptation step (learning rate or convergence factor), and the objective function gradient vector $\nabla(\mathfrak{J}(k))$ is expressed as

$$\nabla(\mathfrak{J}(k)) = \begin{bmatrix} \dfrac{\partial \mathfrak{J}(k)}{\partial \omega_0(k)} \\ \dfrac{\partial \mathfrak{J}(k)}{\partial \omega_1(k)} \\ \text{M} \\ \dfrac{\partial \mathfrak{J}(k)}{\partial \omega_{L-1}(k)} \end{bmatrix} \tag{3.5}$$

where $\partial \mathfrak{J}(k)/\partial \omega_r(k)$ is the partial derivative of the objective function with respect to the r^{th} coefficient $\omega(k)$ [**1**].

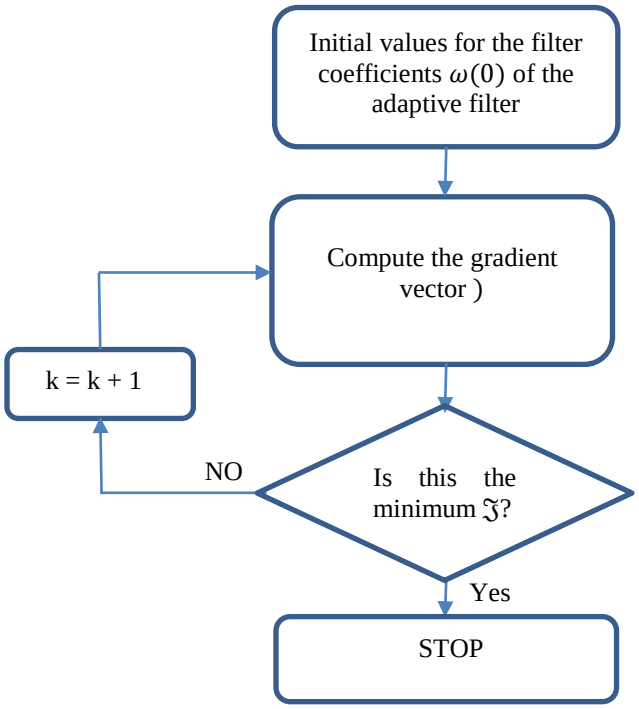


Fig. 3.2 Flow chart explains how to find $\Im_{min}$.

One of the objective functions commonly used is the MSE given by:

$$\Im_k = \varsigma(k) = E\left[e^2(k)\right] = E\left[r^2(k) - 2r(k)y(k) + y^2(k)\right] \quad (3.6)$$

Substituting equation (3.3) in the objective function above, we have:

$$\varsigma(k) = E\left[r^2(k) - 2r(k)\omega^T(n)x(k) + \omega^T(k)x(k)x(k)^T\omega(k)\right]$$

$$= E\left[r^2(k)\right] - 2E\left[r(k)\omega^T(k)x(k)\right] + E\left[\omega^T(k)x(k)x^T(k)\omega(k)\right] \quad (3.7)$$

$$= a - 2b^T p + b^T Rb$$

where $a = E[r^2(k)]$, $b = \omega(k)$, $R = E[x(k)x^T(k)]$ is the autocorrelation matrix of the input signal, and $\mathbf{p} = E[r(k)x(k)]$ is the cross correlation vector between the input

vector and the required signal. The gradient vector for the objective function above is expressed as:

$$\nabla(\Im(k)) = \frac{\partial \varsigma(k)}{\partial \omega(k)} = -2p + 2R\omega(k) \tag{3.8}$$

Substituting this result into equation (3.4) we have the recursive equation based on the gradient descent method given by:

$$\omega(k+1) = \omega(k) + 2\gamma[p - R\omega(k)] \tag{3.9}$$

The optimal solution to equation (3.7) is obtained by $\partial\zeta(k)/\partial\omega(k) = 0$. This is known as the Wiener-Hopf solution [**25**], which is defined as:

$$\omega_0 = R^{-1} p \tag{3.10}$$

However, in practice, accurate estimates for R and $\mathbf{p}$ are not available and the gradient vector must be estimated from the available data. In this sense, two algorithms will be presented below based on statistical gradient: Least-Mean-Square (LMS) and normalised LMS (NLMS).

3.2.3 *Least Mean-Square Algorithm (LMS)*

As it is very difficult to have all of the relevant information to accurately compute the gradient in equation (3.8), the LMS stochastically searches the available input data to reach the average of the optimum solution. If acceptable estimates are obtained of both R and $\mathbf{p}$, denoted as $\tilde{R}(k)$ and $\tilde{\mathbf{p}}(k)$, respectively, a steepest-descent-based algorithm can be employed to look for the Wiener solution of equation (3.10) as explained in the two equations below:

$$\omega(k+1) = \omega(k) - \gamma\tilde{\beta}_{\omega}(k) \tag{3.11}$$

where $\tilde{\beta}_{\omega}(k) = 2\left[\tilde{\mathbf{p}} - \tilde{R}\omega(k)\right]$ represents the estimated gradient vector of the objective function. This estimate is obtained by deriving the instantaneous value of the squared error with respect to the elements of $\omega(k)$, i.e.:

$$\begin{aligned}\widetilde{\beta}_{\omega}(k) &= -2r(k)x(k) + 2x(k)x^{T}(k)\omega(k) \\ &= 2x(k)\left[-r(k) + x^{T}(k)\omega(k)\right] \\ &= -2e(k)x(k),\end{aligned} \tag{3.12}$$

where $x(k)x^T(k)$ and $r(k)x(k)$ represent the estimates of the matrix R and the vector ρ, respectively. Substituting this result into equation (3.11) will have:

$$\omega(k+1) = \omega(k) + 2\gamma e(k)x(k) \tag{3.13}$$

where γ is the learning rate that should be chosen to ensure convergence of the algorithm in stationary states [**95**]. The γ values for algorithm convergence on average is given by the range:

$$0 < \gamma < \frac{1}{\upsilon_{max}} \tag{3.14}$$

where $\nu_{\max}$ is the largest eigenvalue of matrix R.

3.2.3.1 *Excess MSE (EMSE)*

The adaptive filter coefficients converge on average for the optimum ω_0, whereas a deviation occurs instantly $d\omega(k) = \omega(k) - \omega_0$ causing an increase (excess) of the MSE in the case LMS, and given by [**96**]

$$\varsigma = \lim_{n\to\infty} d\zeta(k) \approx \frac{\sigma^2\gamma tr[R]}{1 - \gamma tr[R]} \tag{3.15}$$

For $\gamma tr[R] \ll 1$, the excess MSE can be approximated to:

$$\varsigma \approx \sigma^2\gamma tr[R] = L\gamma\sigma^2\sigma_s^2 \tag{3.16}$$

the length of the adaptive filter is represented by L, and σ_s^2 and $\sigma^2$are the variances of the input signal and additive noise, respectively.

3.2.4 *Normalised LMS Algorithm (NLMS)*

In the LMS, $e(k)x(k)$ correction is applied to the adaptive coefficient vector $\omega(k)$ and directly proportional to the size of the input vector $x(k)$. When these vectors are very large, the LMS suffers from noise amplification in the gradient [**1**]. To overcome this difficulty, an algorithm with a normalised LMS learning rate was proposed. The formula for updating the adaptive filter coefficients is as follows:

$$\omega(k+1) = \omega(k) + \frac{e(k)x(k)}{x^T(k)x(k)} \tag{3.17}$$

In practice, a learning rate, γ_{NLMS}, can be included in the update equation. The parameter ε should also be included to control the learning rate variable when the product $x^T(k)\,x(k)$ becomes very small. Thus, we can rewrite equation (3.17):

$$\omega(k+1) = \omega(k) + \frac{\gamma_{NLMS}}{\varepsilon + x^T(k)x(k)} e(k)x(k) \tag{3.18}$$

For convergence of the NLMS algorithm, the constant learning rate must satisfy the following condition [**14**]

$$0 < \gamma_{NLMS} < 2 \tag{3.19}$$

3.2.5 *Affine Projection (AP) Algorithm*

The filter update formula of the Affine projection (AP) algorithm [**14**] for L parameters, can write be written as:

$$\Delta\omega_L(k) = \omega_L(k) - \omega_L(k-1) \tag{3.20}$$

Minimising the following equation under N constraints

$$\left\|\Delta\omega_L(k)\right\|^2 = \Delta\omega_L^T(k)\Delta\omega_L(k) \tag{3. 21}$$

results in:

$$\omega_L^T(k)x_L(k-i) = r(k-i)\,. \tag{3.22}$$

$x_L(k)$ contains the latest L samples, $i = 0,1,\cdots,N-1,$, N is the projection order, and r(k) is the required signal. The following is the AP update equation given by

$$\omega_L(k) = \omega_L(k-1) + A^T(k)\left(A(k)A^T(k)\right)^{-1} e_N(k) \tag{3.23}$$

where

$$\omega_L(k) = \omega_L(k-1) + A^T(k)\left(A(k)A^T(k)\right)^{-1} e_N(k) \tag{3.24}$$

and $e_N(k)$ is a vector with Nx1 size given by

$$e_N(k) = r_N(k) - A(k)\omega_L(k-1) \tag{3.25}$$

and $r_N(n)$ is the desired signal vector of size Nx1

$$r_N^T(k) = \left(r(k), r(k-1), \Lambda\ , r(k-N+1)\right) \tag{3.26}$$

Equation (3.23) can be generalised to include the whole affine projection family:

$$\omega_L(k) = a_1 + a_2 e_{N\tau}(k) \tag{3.27}$$

With $a_2 = \gamma A_\tau^T(k)\left(A_\tau(k)A_\tau^T(k) + \delta I\right)^{-1}$, $a_1 = \omega_L(k-1-\beta)$, $e_{N\tau} = r_{N\tau} - A_\tau\omega_L(n-1-\beta(N-1))$

$$A_\tau(k) = \left(x_L(k), x_L(k-\tau), \Lambda\ , x_L(k-(N-1)\tau)\right)^T \tag{3.28}$$

and

$$r_{N\tau}^T(k) = \left(r(k), r(k-\tau), \Lambda\ , r(k-(N-1)\tau)\right) \tag{3.29}$$

Despite the satisfactory performance of the AP algorithm, it still has a large computational complexity. An Affine Projection Algorithm (AP algorithm) [**14**] is an alternative to the NLMS algorithm. The former algorithm has a convergence speed higher than that of the NLMS algorithm. The disadvantages of the AP algorithm are its computational complexity and a performance that also degrades in the presence of impulse interference [**37**]. In the AP algorithm, when the convergence is increased with the projection order, there is a simultaneous increase in the computational complexity.

3.2.6 *The Time-Varying LMS Algorithm*

The behaviour of this algorithm is similar to the conventional LMS algorithm; the only difference is the time dependent convergence parameter γ_k. In the case of narrow-band signals, an optimal γ_0 should be selected for the centre frequency f_0. To achieve this goal, the conventional LMS algorithm would be the best choice with a single-tone frequency f_0 to catch the best γ, γ_0 which will update the convergence parameter γ_k according to the formula [**97**]:

$$\gamma_k = \alpha_k \gamma_0 \tag{3.30}$$

where α_k is a decaying factor. The decaying law that is used in [**97**] is:

$$\alpha_k = A^{\frac{1}{1+mk^n}}, \qquad (C, m, \text{and } n) > 0 \tag{3.31}$$

the constants *A*, *m*, *n* determine the norm and the degree of reduction for α_k. Considering equation (3.31), *A* should be greater than one. When *A* is equal to one, α_k is also equal to one and then the time-varying LMS (TV-LMS) will behave as the conventional LMS. The TV-LMS is shown in the following equations [**98**], [**97**].

$$y(k) = \omega(k-1)x^T(k) \tag{3.32}$$

$$e(k) = r(k) - y(k) \tag{3.33}$$

$$\omega(k) = \omega(k-1) + \gamma_k e(k)x(k) \tag{3.34}$$

3.2.7 *The Recursive Least-Square (RLS) Algorithm*

The derivation of the recursive least-square (RLS) [**1**] algorithm is presented as follows:

$$\Im(k) = \sum_{r=1}^{k} \lambda^{k-r} e^2(r) \tag{3.35}$$

where λ is near to or less than 1. The auto- and cross-correlation vectors are written as:

$$R \approx R(k) = \sum_{r=1}^{k} \lambda^{k-r} x(r) x^T(r) \tag{3.36}$$

and $\mathbf{p} \approx \mathbf{p}(n) = \sum_{r=1}^{k} \lambda^{k-r}\, r(k)x(k)$, respectively. Taking the derivation in reference [**1**], the RLS algorithm is expressed as:

$$\omega(k+1) = \omega(k) + g(k)e(k) \tag{3.37}$$

where the updating gain vector is defined as

$$g(k) = \frac{\rho(k)}{1 + x^T(k)\rho(k)} \tag{3.38}$$

and

$$\rho(k) = \lambda^{-1} P(k-1)x(k) \tag{3.39}$$

The inverse correlation matrix of the input $P(k) \equiv R^{-1}(k)$ is calculated as:

$$P(k) = \lambda^{-1} P(k-1) - g(k)\rho^T(k) \tag{3.40}$$

Note that the AP algorithm is getting closer to the RLS algorithm as P increases to M and $\lambda = 1$. There is a strong connection between the NLMS and RLS algorithms as both converge to the same optimum weight vector. The NLMS and AP algorithms are cheaper to implement than the RLS algorithm [**99**].

Chapter 4 discusses the basics of filter banks and multirate systems that will be used to construct our structure, due to their efficient computations and their ability to improve the convergence rate. Chapter 5 will present in detail the implementation of the proposed structure on the system identification.

3.3 Conclusion

We summarise the performances and the limitations of the adaptive algorithms discussed in this chapter and their significance to the problems tackled in our work. The iteration in the gradient method is inexpensive. This method does not need second derivatives. The limitation is in that the convergence speed is highly dependent on the

adaptation step. The LMS is a simple algorithm with low computation and simply implemented. The limitation of the LMS is its sensitivity to nonstationary environments. The performance of the NLMS algorithm drifts in time in applications that have slow changes in signal statistics. The NLMS algorithm convergence is obtained when the MSE is decreased to a minimum value. The time required for the algorithm to converge is inversely proportional to the adaptation size; to avoid this problem, random and large adaption sizes are used to increase the convergence speed due to the stability constraint. In noise-free environments, the NLMS algorithm converges to the minimum MSE and stays there as long as the gradient is zero at the optimum solution. Actually, the NLMS algorithm employs an approximate estimate gradient, which causes an arbitrary update to the coefficients around the optimum values. This produces additional noise at the output in the steady state. Therefore, employing a longer filter not only more expensive, but also presents more noise. To reach a better steady state performance, a smaller adaptation step is needed; however, it leads to a slower convergence. The affine projection algorithm is more accurate than LMS and NLMS. However, the MSE error of the affine projection algorithm is less than the LMS and NLMS, respectively. In general, the affine projection algorithm has a better performance than LMS and NLMS. The expense of the good performance is a higher computational cost. The performance of the time-varying LMS is better than that of the LMS algorithm in terms of the convergence speed. The TV-LMS and the LMS algorithms have less simulation time than the RLS. On the other hand, the TV-LMS has better MSE performance for a higher bandwidth, and this is considered to be a limitation of the TV-LMS algorithm in applications that require larger bandwidth. The advantage of the RLS algorithm is that its convergence speed is higher than the LMS algorithm. On the other hand, the RLS algorithm has a computational complexity higher than the LMS algorithm; thus, a longer computational time will be required in addition to a greater propensity to numerical instability. As the problem of this research is to retrieve the source signals from a convolutive mixture system using an UEPS that requires a faster convergence speed and less computational complexity, we chose the LMS algorithm as it is simpler and has a lower computation cost. The algorithm uses a normalised adaptation step that works at the smallest decimation rate and reduces the errors between bands.

Chapter 4
Filter Banks and Multirate Systems

4.1 Introduction

This chapter presents the concept, components and types of structures used in our proposed works. These topics are important for understanding the forthcoming chapters. Filter banks and multirate systems are used in various applications: audio systems, digital encoding and compression of voice and image signals, adaptive filtering, A/D conversion, sample rate reduction, and multiplexing of signals in digital telephony [**100**].

The concept of a maximally decimated filter bank with EPS is presented in Section 4.2; its performance limitations are discussed; a possible solution for this is presented by using the perfect reconstruction performance. Section 4.6.2 studies different prototypes solutions for the perfect reconstruction filter banks, their performance and limitations are highlighted. Octave bands are discussed in Section 4.3 and used in our work to increase the convergence speed. The non-maximally decimated filter bank is explained in brief in Section 4.2.1, but not used in our first proposed work as it was more appropriate for BSS applications with UEPS, see Chapter 7. The NMD_UEPS introduces less aliasing among all bands. This is an important feature that provides better performance in BSS applications. However, the maximally-decimated filter banks only cancel the aliasing between adjacent bands. This is enough in system identification, but it did not present satisfactory results in BSS. Section 4.4 presents the basic structure of the maximally decimated filter banks of an UEPS. An improvement of this structure is proposed in Chapter 5. Section 4.5 explains the polyphase technique that is used in our design for a perfect reconstruction filter bank that reduces the complexity and eliminates the reconstruction errors. The use of polyphase representation is discussed for EP and UEPS filter bank forms. The chapter concludes with a presentation of different prototypes. The performances and limitations of these prototypes are mentioned and their relevance to the motivation and aims of this work are highlighted.

In recent decades, some multiband adaptive filtering schemes have been presented with the aim of accelerating the convergence of the input signals that are correlated over time (colour signals). In some cases, the multiband adaptive filtering schemes are used to reduce the computational cost by promoting adaptation of the coefficients whose sampling rates are below that of the input signal. However, these schemes have an input-output delay and spectrum overlap between the various bands that should be reduced prior to promoting adaptation of the filters. Lian and Wei [**27**] proposed a scheme with maximum decimation able to make almost an exact modelling of FIR systems. They considered that there is a spectrum overlap between adjacent bands. In this case, both the input signal and the desired signal were decomposed into two bands, and the error generated in each band was used to update the respective adaptive filters related to the band. The idea of this work is used in our proposed method, but with UEPS of multiple bands and the use of a developed normalised adaptation step; the details are discussed in Section 5.4.

In [**101**], two schemes were proposed of non-maximally decimated ($F < L$) filter banks. These adaptive filters were adapted at a rate F times smaller than the input signal; however, as the effect of the overlapping spectrum is directly proportional to the decimation factor, the lower the value of F the smaller the minimum mean square error of the scheme. For fixed values of L and F, one can obtain an optimum filter bank that minimises the mean square error of the final scheme. The difference between the two proposed schemes is that in the first scheme, the desired signal is decomposed; while in the second one, the final error signal of the scheme is decomposed into multiple bands.

Two other schemes have been proposed in [**27**], [**102**]. The first scheme uses an analysis bank without decimation followed by adaptive filters of non-zero coefficients[3], whereas the second one, which is derived from the first, uses a maximally decimated filter bank with perfect reconstruction and adaptive filters operating at a reduced rate.

[3] A non-zero filter is a filter in which most of its coefficients are equal to zero or assume a very small value. The other coefficients are the non-zero coefficients of the filter. The filter coefficients of these filters are updated separately, by setting the adaptation size according to the predicted filter coefficient. This filter is employed to speed up the convergence rate.

Most adaptive multiband schemes employ EP filter banks. Some works [**26**], [**103**], however, have shown that adaptive UEPS with multiple bands can perform better than EPS in terms of convergence speed and/or modelling error due to their flexibility. Through the dynamic allocation of these multiple bands, which avoids the high energy signal components in the vicinity of the multiple bands boundaries [**41**], or by selecting bandwidths and decimation factors that minimise the power of the polyphase components (of orders other than zero) of the unknown system. The use of adaptive filters in UEPS with multiple bands can result in a substantial reduction of the modelling error when compared to the use of the EPS with multiple bands. On the other hand, with some knowledge of the spectral characteristics of the input signal, the convergence rate can be significantly improved using adaptive algorithms in UEPS with multiple bands and a normalised adaptation step [**26**], [**103**]. The UEPS presented in [**103**] employs multiple bands without maximal decimation. In [**26**] a MD_UEPS was derived only with three bands.

In this work, we will extend the results of [**26**] to an arbitrary number of bands, and derive a gradient-type matching algorithm, which works at the lowest sample rate among the involved bands and employs a step of normalisation, resulting in an improved convergence rate for coloured signals when compared to the conventional full-band LMS algorithm and two multiband schemes; the first one is a MD_EPS and the other one is an UEPS. As a consequence of the decomposition in frequency of the input signal in the UEPS, different adaptive filters work at different rates. This leads to some particularities in the adaptation algorithm, such as the use of different error samples and convergence factors, with different normalisations in the equation of updating the coefficients [**104**].

4.2 Maximally Decimated Filter Bank with Equall-Passbands

This type is illustrated in Fig. 4.1. The filter sets $\{P_0(z). \; . \; .P_{L-1}(z)\}$ and $\{Q_0(z). \; . \; . \; Q_{L-1}(z)\}$ are called analysis and synthesis bank, respectively. In this system, the input signal is decomposed into L signals through the analysis filters $P_i(z)$. Signals $x_i(k)$ are then decimated by F_i factors, resulting in decimated signals $x_{a,i}(k)$, and then are

interpolated by a factor of F_i, resulting in the signals $\tilde{x}_{b,i}(k)$. These signals are also at the same sampling rate as the input signal $x(k)$ and are recombined by synthesis filters $Q_i(z)$. The analysis filters $P_i(z)$ are intended to decompose the signal $x(k)$ into L different frequency band levels, while the synthesis filters $Q_i(z)$ combine the signals from the L bands in a single signal $\tilde{x}(k)$. Filter bank is called maximally or critically decimated [**100**] when

$$\sum_{i=0}^{L-1}\left(\frac{1}{F_i}\right)=1. \tag{4.1}$$

In the reconstruction output, due to the processing stage, there exists an error of filter bank reconstruction that may have different characteristics: amplitude distortion, phase distortion and overlapping of the spectrum (aliasing) between the frequency-responses in $P_i(z)$ and $Q_i(z)$. Fortunately, these errors are avoided in the maximally decimated filter bank design, leading to a perfect reconstruction of the input signal [**100**].

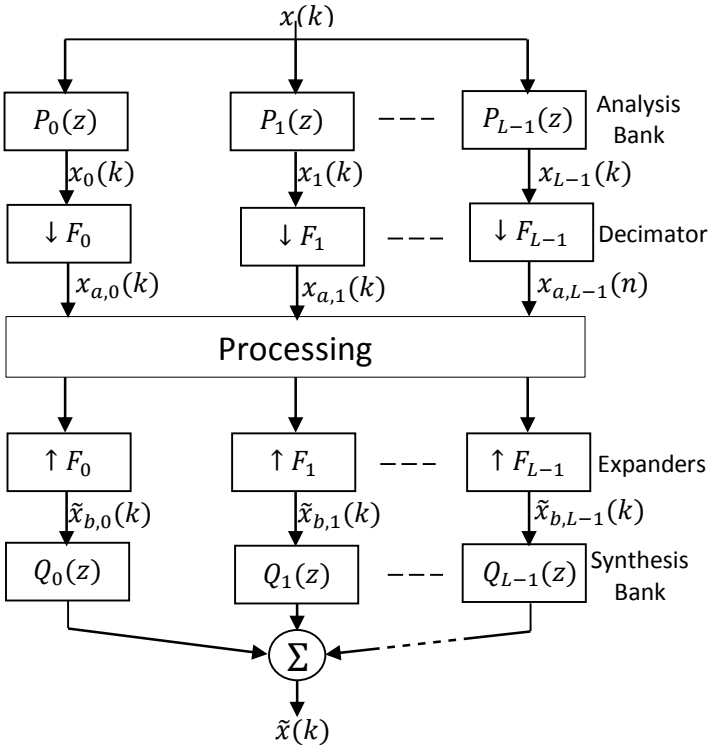


Fig. 4.1 Filter Bank with EP

Usually, $\tilde{x}(k)$ represents the recovered shifted input, D is the delay introduced by the filter bank and the factor $1/k$ is the resulting decimation process, then $\tilde{x}(k) = \frac{1}{k}x(n-D)$. However, the latter can be compensated by a proper design of the analysis and synthesis filters. Section 4.5.1 discusses the filter bank with EP using the polyphase representation.

Historically, the theory of Perfect Reconstruction (PR) began with the QMF Quadrature Mirror Filter (QMF) filter bank type [**105**], but only in [**106**] and [**107**] it was shown how to eliminate the three types of distortion, mentioned above, by using a filter bank with two bands (L = 2).

In fact, we use the mentioned advantages of the EP filter bank by using two bands to construct an "UEPS tree" [**108**]. This structure is also known as octave filter banks (Section 4.3), which shows some advantages in signals that have more energy concentrated in low frequencies, as audio and voice signals. An improvement to this structure is proposed in Section 5.2. We choose the MD_UEPS to cancel the aliasing between neighbouring bands, which is enough in the system identification problem, but did not present satisfactory results in BSS.

4.2.1 *Non-Maximally Decimated Filter Bank*

This type of filter bank is explained in this chapter in order to clearly distinguish it from the maximally decimated filter band, mentioned earlier; moreover it provides the opportunity to justify the use of non-maximally decimated filter bank as a part of our proposed structure in system identification application.

Consider L bands and the decimation factor F_i in each band. A filter bank is called a non-maximally decimated filter bank when $\sum_{i=0}^{L-1}\frac{1}{F_i} > 1$. Fig. 4.2 [**108**] shows a bank of 4 subsampled bands, decimated and expanded by a factor F_i = 2 for i = 0, 1, 2 and 3 (in this case, $\sum_{i=0}^{3}\frac{1}{F_i}$ =2).

We chose to use a non-maximally decimated filtering with multiple bands in BSS applications because it has a smaller spectrum overlapping [**108**], in other words, less aliasing among all bands in our proposed UEPS. This important feature provides a better performance in BSS applications. An in depth discussion is provided in Chapter 7.

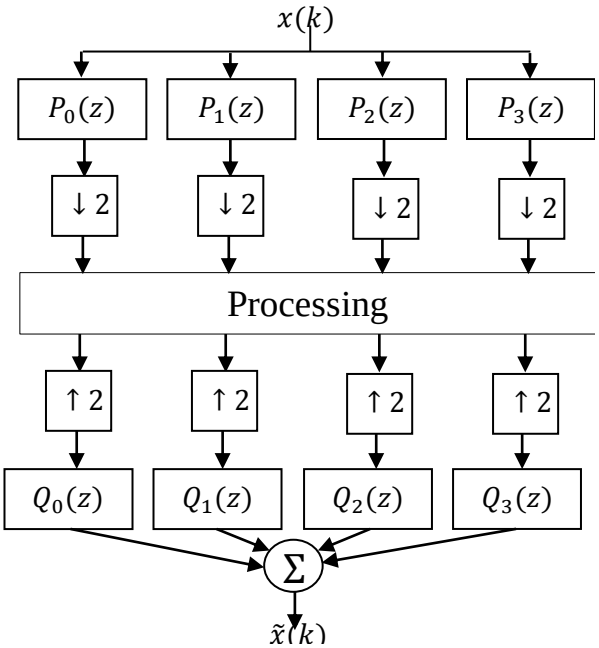


Fig. 4.2 Example of a NMD filter bank.

4.3 Tree-Structured Filter Bank

This section discusses the important part that is used to deal with the large energy that is concentrated in the speech signal by adding narrow octave bands. This is done by breaking down the signal at the input into many narrow bands to increase the speed of convergence.

The input signal of this filter bank is divided into two bands by a maximally decimated analysis filter bank; then, the signal of each band is once again divided and decimated.

Each division and decimation stage is considered as a level of decomposition. The higher the level number, the greater is the number of bands in the filter bank. These bands

are then recombined into pairs through a two-band synthesis filter bank. This is a MD_EPS system with two levels of decomposition as shown in Fig. 4.3 (a), also known as a binary-tree structured filter bank [**109**]. Filters [$P^{0,0}(z)$, $P^{1,0}(z)$] and [$Q^{0,0}(z)$ and $Q^{1,0}(z)$] construct the analysis and synthesis filter bank, respectively, with $L = 2$ bands, where the first superscript index is equal to 0 and represents a lowpass filter (LPF) and 1 represents a high pass filter (HPF), while the second superscript index indicates the level of decomposition of the tree structure. Using noble identities [**108**], the decimators of the r^{th} level of the analysis bank will be moved down the filters $P^{0,r+1}(z)$ and $P^{1,r+1}(z)$ of the next level. Similarly, we can move the expanders r^{th} level of the synthesis bank up to the filters $Q^{0,r+1}(z)$ and $Q^{1,\,r+1}(z)$ of the previous level. We can then obtain the equivalent representation for the binary filter bank shown in (b), where $L = 4$ analysis filters $P_i(z)$ that are given by

$$\begin{aligned} P_0(z) &= P^{0,0}(z)P^{0,1}\left(z^3\right),\ P_1(z) = P^{0,0}(z)P^{1,1}\left(z^2\right), \\ P_2(z) &= P^{1,0}(z)P^{0,1}\left(z^2\right),\quad P_3(z) = P^{1,0}(z)P^{1,1}\left(z^2\right), \end{aligned} \tag{4.2}$$

and synthesis filters $Q_i(z)$ are written as

$$\begin{aligned} Q_0(z) &= Q^{0,0}(z)Q^{0,1}\left(z^2\right),\ Q_1(z) = Q^{0,0}(z)Q^{1,1}\left(z^2\right), \\ Q_2(z) &= Q^{1,0}(z)Q^{0,1}\left(z^2\right),\ Q_3(z) = Q^{1,0}(z)Q^{1,1}\left(z^2\right) \end{aligned} \tag{4.3}$$

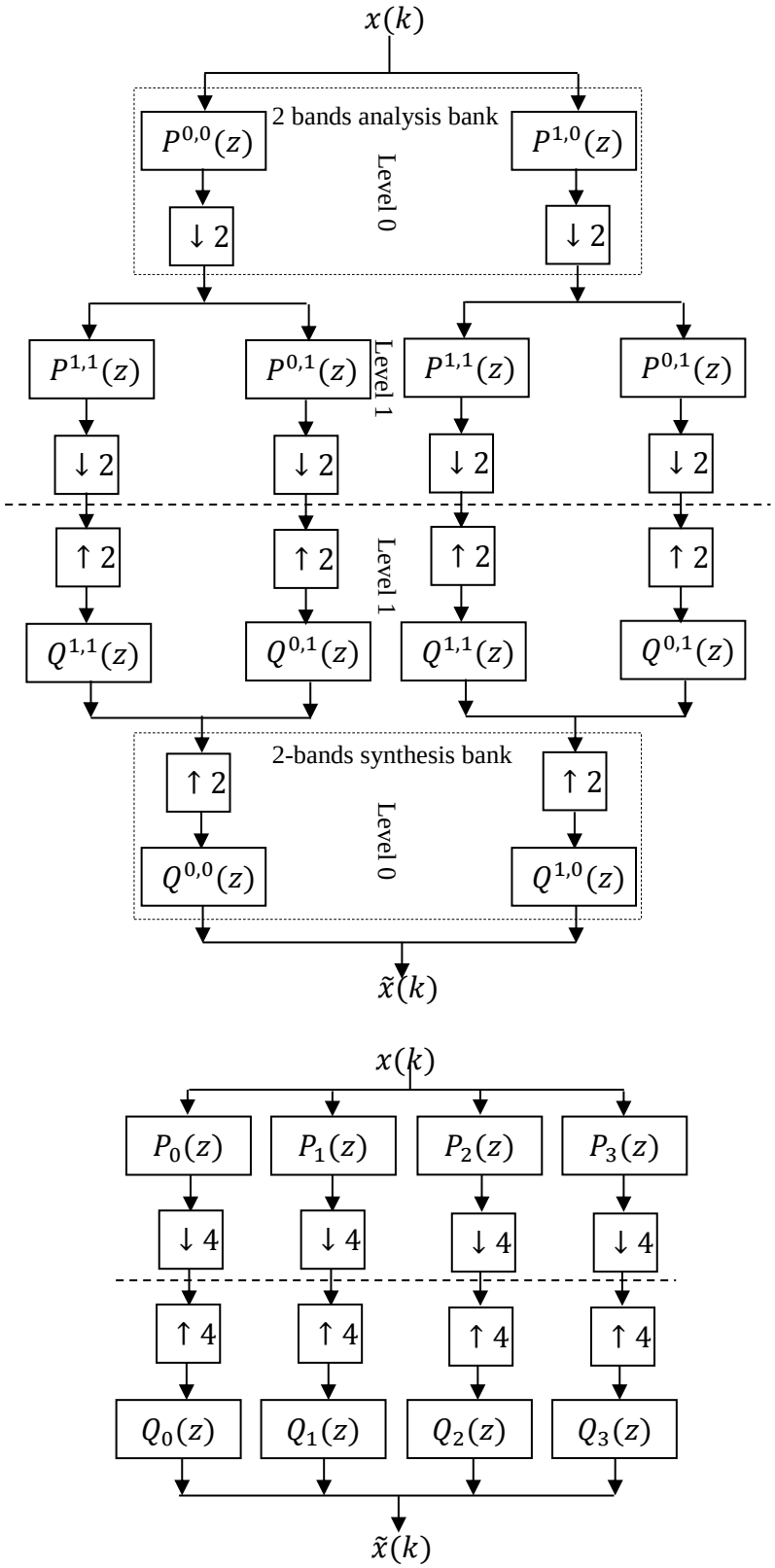


Fig. 4.3 (a) Example of a maximally decimated tree-structured filter bank and (b) its equivalent representation.

4.4 The basic principle of the MD_UEPS

This section presents the basic structure of MD_UEPS. This is the structure that we have improved and proposed in Chapter 5. The filter bank that divides the input signal into L bands consisting of different passbands is called a filter bank with UEPS. This type of filter bank can be obtained in various ways, for example, from a filter bank with EP in which the signals of each band are combined to produce a new UEPS. Another way is based on the decomposition view, given in Section 4.3 [**99**]. Fig. 4.4(a) shows a binary tree filter bank with UEPS that uses $L = 4$, obtained by a tree decomposition of a 2-band MD_EP filter bank. In this case, three levels are needed for decomposition, where each subsequent level is obtained by decomposing the decimated signal at the output of the low-pass filter in the previous stage ($P^{0,r}(z)$). The L analysis filters $P_i(z)$ of equivalent representation is given by

$$\begin{aligned} P_0(z) &= \prod_{c=0}^{L-2} P^{0,c}\left(z^{2}\right) \\ P_i(z) &= P^{1,L-1-i}\left(z^{2^{L-1-i}}\right)\prod_{c=0}^{L-i-2} P^{0,c}\left(z^{2^{c}}\right) \end{aligned} \tag{4.4}$$

where $Q_i(z)$ are the synthesis filters, given by

$$\begin{aligned} Q_0(z) &= \prod_{c=0}^{L-2} Q^{0,c}\left(z^{2^{c}}\right) \\ Q_i(z) &= Q^{1,L-1-i}\left(z^{2^{L-1-i}}\right)\prod_{c=0}^{L-i-2} Q^{0,c}\left(z^{2^{c}}\right) \end{aligned} \tag{4.5}$$

where i =1, …, L-1.

Delays d_i (will be further explained in Section 5.2) are needed to compensate the difference in the length between the analysis filters. The decimation and expansion factors of the equivalent representation are obtained as follows:

$$\begin{cases} F_0 = 2^{L-1}, \\ F_i = 2^{L-i}, \quad for \quad 1 \le i < L-1. \end{cases} . \tag{4.6}$$

Fig. 4.5 illustrates the frequency response of the analysis and synthesis filters of the filter bank with UEPS shown in Fig. 4.4(b). In this type, each increment of a level of decomposition implies a reduction of the passband width to half of the corresponding width of the previous level. Therefore, this type of structure is also known as a filter bank with octave frequency decomposition, or simply an octave filter bank. For more efficient implementation, Section 4.5.14.5.2 discusses the filter bank with EP using the polyphase representation [**99**].

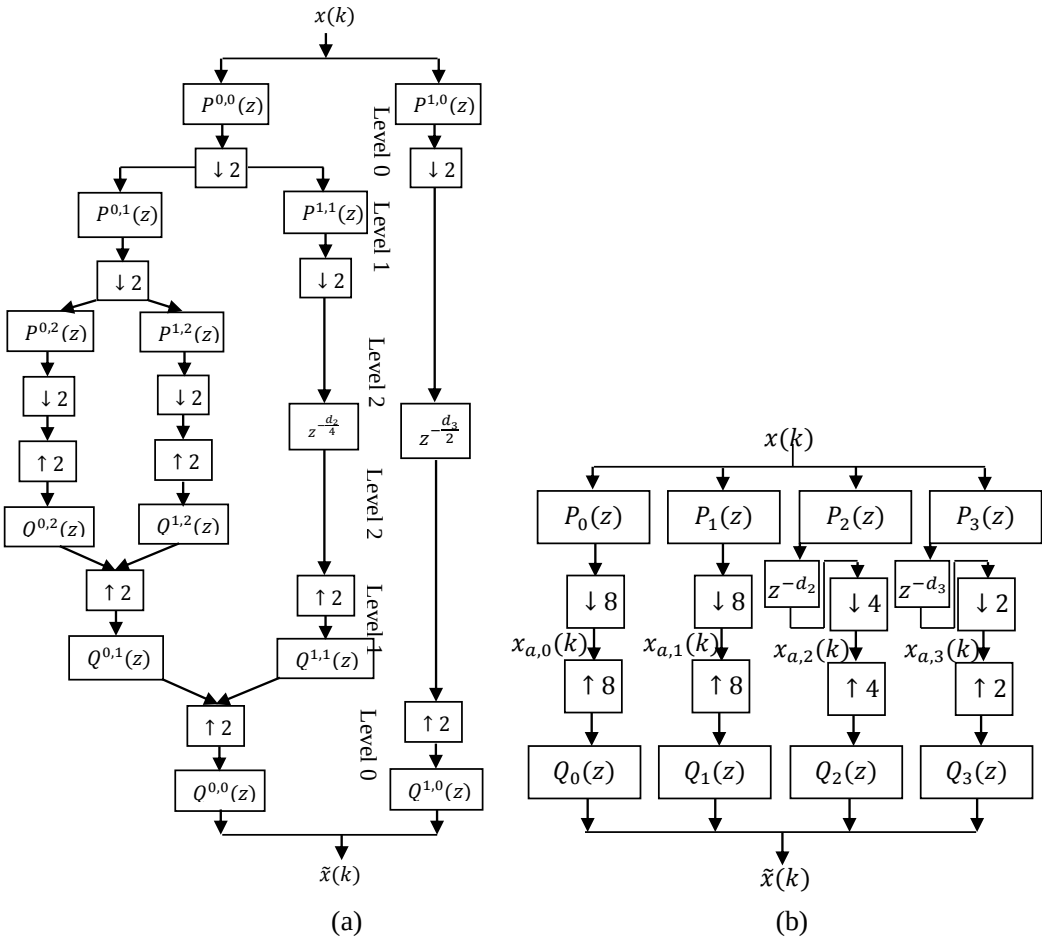


Fig. 4.4: (a) maximally decimated binary tree filter bank with UEPS and (b) equivalent representation.

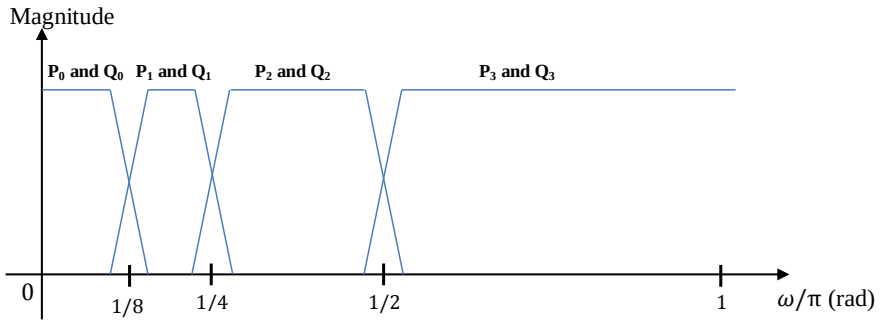


Fig. 4.5: Frequency Response of an octave filter bank with L = 4.

4.5 Polyphase Representation

The polyphase representation provides efficient implementation as it is used with cosine-modulated filter banks (Section 4.6) that is known with their real-valued nature. This section presents its use with EP and UEPS filter banks.

A major improvement in multirate signal processing is attributed to the polyphase representation of analysis and synthesis banks. The representation made it possible to simplify the theoretical analysis of multirate systems and provides computation efficiency in implementing both decimators and interpolators as in the filter banks [**108**], [**110**] .

4.5.1 *Filter Bank with EP*

Fig. 4.6 shows the general structure of a multirate system with EP using polyphase representation, assuming that the decimation and interpolation factors are all equal to F. Decomposing the analysis filters $P_r(z)$ as follows:

$$P_r(z) = \sum_{c=0}^{F-1} z^{-c} P_{r,c}\left(z^F\right) \tag{4.7}$$

where $P_{r,c}(z)$ are polyphase components of a type-I[4] of the r^{th} analysis filter $p_r(n)$, i.e.,

$$P_{r,c}(z) = \sum_{\lambda=-\infty}^{\infty} p_r(\lambda F + c) z^{-\lambda} \tag{4.8}$$

the analysis polyphase matrix $P_m(z^F)$ of dimension $L \times F$ is defined as

$$P_m\left(z^F\right) = \begin{bmatrix} P_{0,0}\left(z^F\right) & P_{0,1}\left(z^F\right) & \Lambda & P_{0,F-1}\left(z^F\right) \\ P_{1,0}\left(z^F\right) & P_{1,1}\left(z^F\right) & \Lambda & P_{1,F-1}\left(z^F\right) \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ P_{L-1,0}\left(z^F\right) & P_{L-1,1}\left(z^F\right) & \Lambda & P_{L-1,F-1}\left(z^F\right) \end{bmatrix} \tag{4.9}$$

[4] There are two types of polyphase decomposition: type I and type II. The former is associated with the analysis bank and the latter with the synthesis bank [**108**], [**119**].

The polyphase decomposition representation of the synthesis filter bank $Q_r(z)$ of type-II [**111**]can be written as:

$$Q_r(z) = \sum_{c=0}^{F-1} z^{-(F-1-c)} Q_{r,c}\left(z^F\right) \tag{4.10}$$

where $F_{r,c}(z)$are polyphase components of the r^{th} synthesis filter $q_r(k)$, that is,

$$Q_{r,c}(z) = \sum_{\lambda=-\infty}^{\infty} q_r\left(F(\lambda+1)-c-1\right) z^{-\lambda} \tag{4.11}$$

.

The synthesis polyphase matrix $Q_m(z^F)$ of dimension $F \times L$ is defined as

$$P_m\left(z^F\right) = \begin{bmatrix} Q_{0,0}\left(z^F\right) & Q_{1,0}\left(z^F\right) & \Lambda & Q_{L-1,0}\left(z^F\right) \\ Q_{0,1}\left(z^F\right) & Q_{1,1}\left(z^F\right) & \Lambda & Q_{L-1,1}\left(z^F\right) \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ Q_{0,F-1}\left(z^F\right) & Q_{1,F-1}\left(z^F\right) & \Lambda & Q_{L-1,F-1}\left(z^F\right) \end{bmatrix} , \tag{4.12}$$

Using noble identities [**111**], the scheme of Fig. 4.6 can be simplified, as shown in Fig. 4.7, where the analysis and synthesis filters operate at a rate F times smaller than the input signal $x(k)$.

For a multirate system of L bands, the conditions to achieve perfect reconstruction were developed in [**112**] and [**113**] for an orthogonal[5] filter bank, using the paraunitarity property of the polyphase matrix [**110**], that is;

$$Q_m(z)P_m(z) = z^{-D_0} I \ , \tag{4.13}$$

[5] A filter bank is said to be orthogonal when its polyphase matrices are paraunitary. In the time domain, the condition of paraunitarity is satisfied when the coefficients of the synthesis filters are time-reversed, with respect to the coefficients of the analysis filters [**119**].

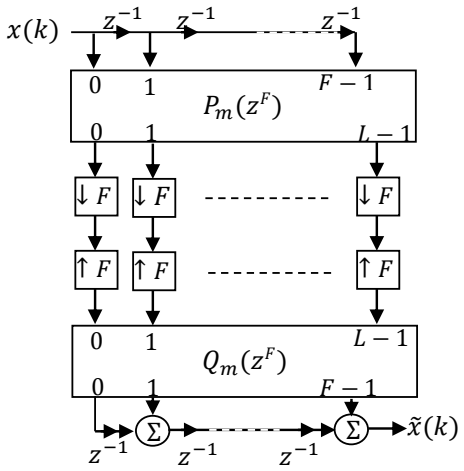


Fig. 4.6 Scheme of the general multirate polyphase representation with the analysis and synthesis banks.

where $Q_m(z) = P_m^T(z^{-1})$ and D_0 is the delay introduced in the polyphase domain; the total system delay is $D = F - 1 + D_0F$.

Since the design of analysis and synthesis filters for L bands still appeared to be very complex, an alternative solution was to obtain them from a single filter [**114**], [**115**], [**116**].

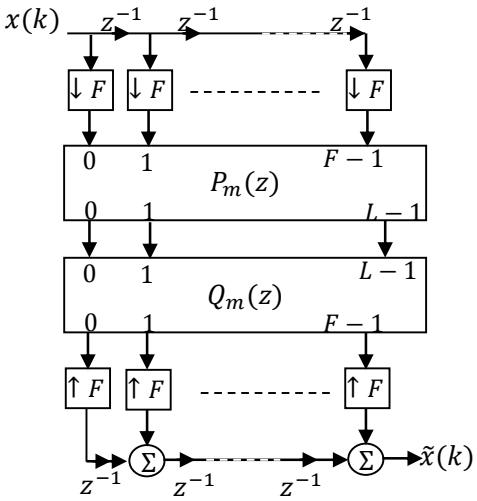


Fig. 4.7: Simplified polyphase Filter

4.5.1.1 ***MD_EPS***

The MD_EPS method [**42**] uses a filter bank with non-zero filters [**27**]. Some restrictions are added to make this structure able to model FIR. Filter bank configurations with maximally decimated signals at each band were attained from the non-zero adaptive filters. To achieve such configurations only L filters is adapted in L-band scheme. The performance of the adaptation approach is evaluated by studying the convergence rate for coloured inputs and high order filters. This approach has a low computational cost for high order filters. Our proposed MD_UEPS will be compared with this method, see Fig. 5.10.

4.5.2 ***Filter Bank of UEPS***

The basic idea is to expand the filter bank of UEPS with L bands, different decimation factors F_i in a filter bank with F_0 bands and a single decimation factor K for all bands, given by the least common multiple (LCM) between different factors F_i. For the octave filter bank, shown in Section 4.4, the decimation and expansion factors for all bands are related by:

$$K = LCM(F_0, F_1, \Lambda\ , F_{L-1}) = F_0 \tag{4.14}$$

To achieve this goal, an i^{th} band analysis/synthesis of the tree structure with tree PR of UEPS is shown in Fig. 4.8(a) and then extended into $\ell_i = F_0/F_i$ bands as shown in Fig. 4.8(b).

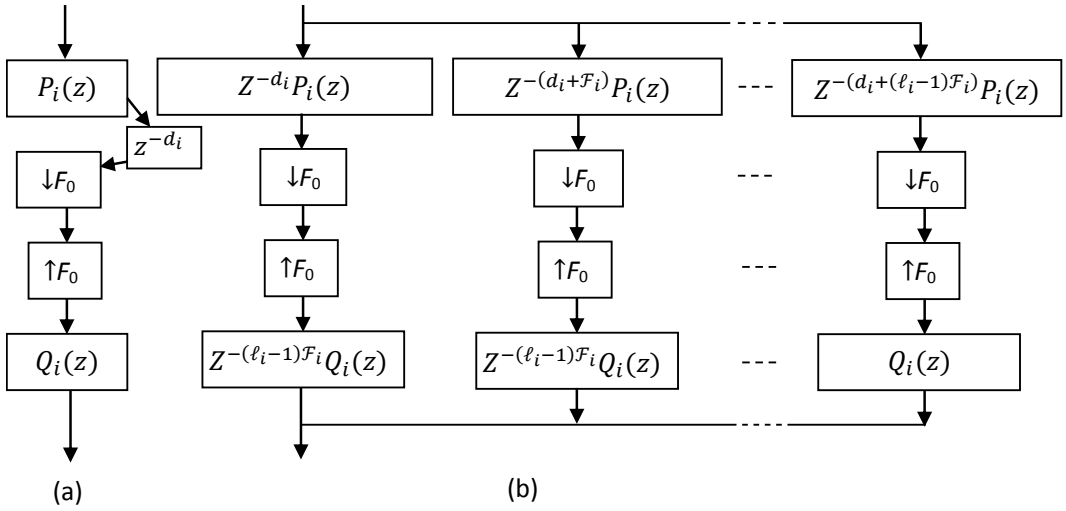


Fig. 4.8 (a) i^{th} band filter bank with UEPS and PR (b) expansion of the i^{th} band into ℓ_i bands.

4.5.2.1 ***NMD_UEPS***

This method proposes a significant analysis for UEPS filters by connecting the filter bank coefficients to the Mean Square Error of each band. The MSE at each band is minimised when the number of filters is reduced and when the decimation factor is minimized. By cautiously selecting a decimation factor, the non-zero polyphase element power is reduced and so the mean square error. Then, by reducing the MSE of each band separately, the overall error will be reduced. Moreover, the bandwidth of each band is selected in such a way to increase the decimation factor of each band. Therefore, this approach is optimal. The computation cost is low as the algorithm does not run continuously. This method also eliminates the abrupt changes produced by altering the

filter bank. The drawback of this method is the slow convergence that may appear with oversampling during the change between the filter banks [**43**]. Our proposed MD_UEPS will be compared with this method, see Fig. 5.10.

4.6 Cosine-Modulated Filter Bank

There are different methods in the literature for designing and implementing filter banks [**108**], [**116**]. This section describes the Cosine Modulated Filter Bank (CMFB).

In these systems, all the *L* analysis and *L* synthesis filters are obtained using cosine modulation of a single prototype filter. Some significant advantages for this system are:

- Computational cost for implementing the analysis and synthesis filter banks is the same as that for a single DCT (Discrete Cosine Transform) filter;
- The required number of parameters to be optimised, in designing the filter bank, is reduced because only the prototype filter has to be optimised;

Assuming that the impulse response is $h(k)$ and the length of the prototype filter is K_{pr}, the analysis and synthesis filter banks are found as follows:

$$p_i(n) = 2h(k)\cos\left[\frac{\pi}{F}(i+0.5)\left(k-\frac{D}{2}\right)+\varphi_i\right], \tag{4.15}$$

$$q_i(n) = 2h(k)\cos\left[\frac{\pi}{F}(i+0.5)\left(k-\frac{D}{2}\right)-\varphi_i\right], \tag{4.16}$$

where $D = K_{pr} - 1$ and $\varphi_i = (-1)^i \frac{\pi}{4}$, for $0 \le i \le L-1$ and $0 \le k \le K_{pr}-1$.

There are different types of CMFBs, among them:

a. Pseudo Quadrature Mirror Filter (PQMF), which is characterised by cancelling the overlap between adjacent spectrum bands [**117**];

b. Near Perfect Reconstruction (NPR) filter, which is characterised by cancelling errors and distortion of the spectrum overlap between adjacent bands [**110**];

c. A Filter bank with Perfect Reconstruction (PR), which is characterised by not having errors of distortion and overlapping [**108**].

4.6.1 *Errors and Aliasing Distortion*

For a filter bank with L bands and $\mathcal{F}$ decimation factors, the relationship between the input and the output signal is [**108**]:

$$\widetilde{X}(z) = \sum_{\lambda=0}^{F-1} T_{\lambda}(z) X\left(z W_F^{\lambda}\right) \tag{4.17}$$

where

$$T_{\lambda}(z) = \frac{1}{F} \sum_{i=0}^{L-1} P_i\left(z W_F^{\lambda}\right) Q_i(z) \tag{4.18}$$

and $W_F = e^{-j\frac{2\pi}{F}}$.

The transfer function $T_0(z)$ multiplies the original spectrum of the input signal and is known as the distortion function. The transfer functions $T_1(z), \cdots, T_{F-1}(z)$ multiplied by the shifted versions of the input signal spectrum; they are known as transfer the functions of an overlapping spectrum. The objective is to design filters to ensure the perfect reconstruction of the input signal; this requires that:

$$\begin{cases} T_0(z) = z^{-D}, \\ \\ T_{\lambda}(z) = 0, \quad for \quad 1 \leq \lambda \leq F-1. \end{cases} \tag{4.19}$$

where the overlap between adjacent bands of the spectrum is represented by $T_1(z)$ and between non-adjacent bands are represented by $T_2(z)$,. . . , $T_{\mathcal{F}-1}(z)$.

4.6.2 *Prototypes for Cosine-Modulated Filter Bank*

In this section, we will present four types of prototype filters: Pseudo-QMF (PQMF), with near-perfect reconstruction (NPR), maximally decimated with perfect reconstruction (PR) and non-maximal decimation to PR, from which the cosine-modulated analysis and synthesis banks will be implemented (equations (4.15) and (4.16)). Fig. 4.9 (a) and (b) show the frequency response of the prototype filter and analysis bank, respectively.

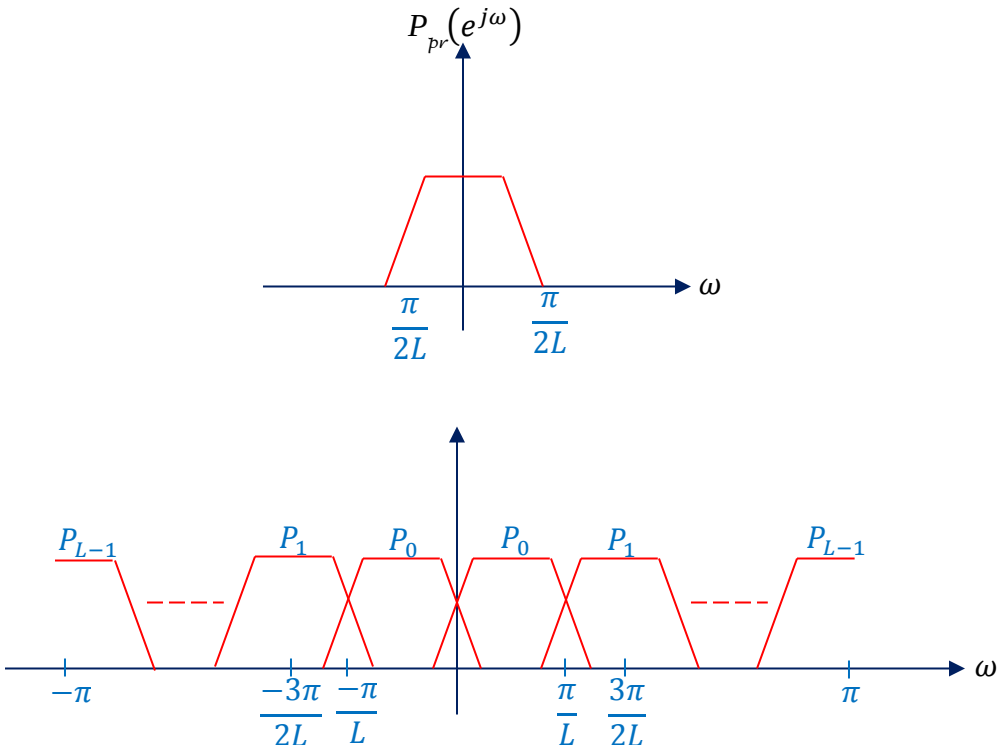


Fig. 4.9 (a) Frequency responses of the prototype filter, and (b) the analysis filter.

4.6.2.1 *Pseudo Quadrature Mirror Filter (PQMF) Prototype*

Initially, techniques were developed for approximate reconstruction systems. These systems are called Pseudo Quadrature Mirror Filter (PQMF) initially studied in [**117**]. In such systems, the analysis $P_i(z)$ and synthesis $Q_i(z)$ filters are selected so that only the spectrum overlap between adjacent bands is cancelled and the distortion function $T_0(z)$ is the only delay. These systems are acceptable in some applications such as speech, image and video processing [**118**]. For the filter bank to have a sufficient flat response over the full range of the frequencies [0, π], it is necessary to force the power complementary of the prototype through the objective function:

$$\varsigma_1 = \sum_{\omega=0}^{\frac{\pi}{L}} \left(\left| P_{pr}\left(e^{j\omega}\right) \right|^2 + \left| P_{pr}\left(e^{j\left(\omega-\frac{\pi}{L}\right)} \right) \right|^2 - 1 \right)^2 \tag{4.20}$$

In order for the filter bank to have a good selectivity and to reduce the overlap of the spectrum between non-adjacent bands, it is important that the prototype filter has a large attenuation in the rejection band. So its energy can be minimised in the rejection range using the following objective function [**118**]:

$$\varsigma_2 = \sum_{\omega=\frac{\pi}{L}+\nu}^{\pi} \left(\left| P_{pr}\left(e^{j\omega}\right) \right| \right)^2 \tag{4.21}$$

where ν controls the selectivity of the prototype.

Finally, the coefficients $p_{pr}(n)$ in $P_{pr}(z)$ that composes the objective function can be optimised by combining the above functions as follows [**118**]:

$$\varsigma = A_1\varsigma_1 + A_2\varsigma_2 \tag{4.22}$$

The prototype $P_{pr}(z)$ is a linear phase FIR filter with real coefficients and the cutoff frequency is $\pi/2L$ (Fig. 4.9). The phase error and spectral overlap between adjacent channels are cancelled, making $q_i(k) = p_{pr_i}(k)(K_{pr} - k)$ and $\varphi_i = (-1)^i \frac{\pi}{4}$ in equation (4.15) [**118**].

To prevent the mentioned aliasing, extra adaptive cross filters are added between the bands. These cross filters will increase the computational cost and decrease the speed of the convergence. This extension is considered a limitation to our objectives. Our goal is to increase the convergence rate and decrease the computational cost that cannot happen with this type of prototypes [**118**].

4.6.2.2 *Near-perfect reconstruction prototype*

A prototype filter $P_{pr}^{(z)}$ of length $K=2(nL + n_1)$ for all n and $0 \leq n_1 \leq L-1$, resulting in a near-perfect reconstruction filter bank with a cosine-modulated input signal, that can be obtained by a spectral decomposition of a $2L$ filter bands with a linear phase [**110**]. This prototype performs well as the reconstruction errors are cancelled and the overlap error of the spectrum at the output of the filter bank is comparable to the attenuation of this prototype in the rejection band. Using the above notation, the $2L$ filter band is defined as follows:

$$S(z) = \sum_{k=0}^{4nL+4n_{11}-2} s(k)z^k \tag{4.23}$$

where

$$s(k) = \begin{cases} 0, & for \quad k = 2(nL + n_1) - 1 - 2\lambda L \quad and \quad \begin{cases} 1 \leq \lambda \leq n-1, & n_1 = 0 \\ 1 \leq \lambda \leq n, & n_1 \neq 0 \end{cases} \\ \dfrac{1}{2L}, & for \quad k = 2(nL + n_1) - 1 \end{cases} \tag{4.24}$$

In [**110**], quadratic constraints were obtained to ensure that the prototype filter corresponds to the spectral decomposition filter $S(z)$ in $2L$ bands, making it possible to

obtain prototypes with near-perfect reconstruction through an optimisation with quadratic constraints. Considering only the case where L is even, the following restrictions to prototype should be imposed:

$$\begin{cases} p_{pr}^{T} D_k p_{pr} = 0, & for \quad \begin{cases} \\ \\ \end{cases} \\ p_{pr}^{T}\left(D_k + \Im D_{k-nL-n_1} + D_{k-nL-n_1}\Im\right)p_{pr} = 0 & for \quad n \le \lambda \le \left[\frac{n+1}{2}\right] - 1 \\ p_{pr}^{T}\left(\Im D_{nL-n_1-1} + D_{nL-n_1-1}\Im\right)p_{pr} = \frac{1}{2L} & \end{cases} \tag{4.25}$$

where

$$\Im = \begin{pmatrix} 0 & \Lambda & 1 \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 1 & \Lambda & 0 \end{pmatrix} \tag{4.26}$$

$$\left[D_k\right]_{r,c} = \begin{cases} 1, & for \quad k = r + c \\ 0, & for \quad any\ k. \end{cases} \tag{4.27}$$

and $(nL+n_1)\times(n_L+n_1)$ is the dimensions of the matrices $\Im$ and D_k, $k = 2L(n\text{-}\ell) + 2n_1 - 1$), and p_{pr} is a vector that consists of the first $nL + n_1$ coefficients in $p_{pr}(k)$, so

$$p_{pr} = [p_{pr}(0) \quad p_{pr}(1) \quad \cdots \quad p_{pr}(nL - n_1 - 1)]^T \tag{4.28}$$

It is necessary for the prototype filter to optimise its energy in the rejection range, according to the following objective function, meeting the n restrictions imposed in equation (4.25):

$$\varsigma = \sum_{\omega=\omega_u}^{\pi}\left(\left|P_{pr}\left(e^{j\omega}\right)\right|\right)^2 \tag{4.29}$$

where $\omega_u = \frac{\pi}{L}$ (corresponding frequency to the start of the rejection range of the prototype filter).

The limitation of this type is in that the prototype filter output is given with a fewer multiplications, at the cost of increased delay. We avoid using this type, as this limitation will be an impediment to our design, and thus contrary to our aspiration to construct a robust structure with minimum delay [**118**].

4.6.2.3 *Maximally Decimated Perfect-Reconstruction Prototype*

This prototype was obtained in [**114**]. The conditions for 2*L* polyphase elements of a prototype filter of length $K = 2nL$, ($n \geq 1$) and linear phase, such that the polyphase matrix of a filter bank is cosine modulated paraunitary [**119**]. A theorem in this reference states: $p(z)$ matrix is the analysis filter bank obtained from the real coefficient, linear phase prototype filter $P_{pr}(z)$ of length $K=2nL$, then, the matrix of polyphase components of $p(z)$ is paraunitary (lossless) if and only if:

a) When *L* even:

$$S_i(z^{-1})S_i(z) + S_{L+i}(z^{-1})S_{L+i}(z) = \frac{1}{2L} \tag{4.30}$$

where $S_i(z)$ are the polyphase components of type-I in $P_{pr}(z)$ for $0 \leq i \leq \frac{L}{2} - 1$.

b) When *L* odd:

$$S_i(z^{-1})S_i(z) + S_{L+i}(z^{-1})S_{L+i}(z) = \frac{1}{2L} \tag{4.31}$$

and

$$2S_{\frac{L-1}{2}}(z^{-1})S_{\frac{L-1}{2}}(z) = \frac{1}{2L} \tag{4.32}$$

for $0 \leq i \leq \left[\frac{L}{2}\right] - 1$, where $[x]$ is the largest integer less than x.

In [**35**], the restrictions above (equations (4.30), (4.31) and (4.32)) were rewritten in a function of the prototype filter coefficients ($P_{pr}(z)$), making it possible to obtain prototypes with perfect reconstruction (PR) through an optimisation with quadratic

constraints. Considering only the case where L is even, we can rewrite equation (4.30) as follows:

$$p_{pr}^{T}\left[\eta_i \Im D_k \eta_i^{T} + \eta_{L+i} \Im D_k \eta_{L+i}^{T}\right] p_{pr} = \begin{cases} 0, & 0 \le k \le n-2 \\ \dfrac{1}{2L}, & k = n-1 \end{cases} \tag{4.33}$$

where

$$p_{pr} = \left[p_{pr}(0) \quad p_{pr}(1) \quad \Lambda \quad p_{pr}(nL-1)\right]^{T} \tag{4.34}$$

and

$$\left[\eta_i\right]_{r,c} = \begin{cases} 1, \begin{cases} r = i + 2cL, & i + 2jL < n\,L \\ r = 2L(l-j) - 1 - i, & i + 2jL \ge n\,L \end{cases} \\ 0, \quad for \quad any \quad r \end{cases} \tag{4.35}$$

where $0 \le i \le \frac{L}{2} - 1$, $0 \le k \le l - 1$, and matrices $\Im$ and D_k are obtained in accordance with equations (4.26) and (4.27), respectively. Taking into account the symmetry of the prototype filter, the $\frac{L}{2}$ conditions of perfect reconstruction in equation (4.30) are rewritten as $\frac{nL}{2}$ quadratic constraints in equation (4.33). Note that the dimensions of p_{pr}, $\boldsymbol{\eta}_i$, $\Im$ and D_k, are ($nL \times 1$), ($nL \times l$), ($n \times n$) and ($n \times n$), respectively. For the prototype filter, it is necessary to optimise its energy in the rejection range, according to the objective function shown in equation (4.29), meeting the $\frac{nL}{2}$ restrictions imposed in equation (4.33).

This type has good stopband attenuations. The approximate error of this type is affordable and can be ignored in audio applications. The complexity is low. The only limitation is in the restrictions of the small number of coefficients. This limitation can be taken into account in the design trade off [**118**]. For all of the above reasons, we will

consider this type in our design of the first proposed work of the UEPS with multiple bands.

4.6.2.4 *Non-maximally decimated perfect-reconstruction Prototypes*

The necessary conditions for PR in CMFB with arbitrary delay D, was developed by Kliewer and Mertins in [**120**]. Considering P_{pr}(z) prototype filter of length $K = 2nL$, and only the case where L is even, we can write the PR conditions as follows:

$$\sum_{l=0}^{2R-1} P_{pr_{i+\lambda F}}(z) P_{pr_{2L-1-i-\lambda F}}(z) = \frac{z^{-D_1}}{2F} \tag{4.36}$$

where $i = 0, \cdots, \left\lceil \frac{F}{2} \right\rceil - 1$, and R is the ratio of the number of L bands and the factor of decimation F, P_{pr_c}(z) are the polyphase components of type-I of the prototype filter, (D_1) is the delay introduced by the analysis-synthesis system and $\lceil x \rceil$ is the smallest integer greater than x. The delay in the polyphase domain D_0 is defined as:

$$D_0 = 2RD_1 + 2R - 1 \tag{4.37}$$

where D_1 may vary from 0 to $2n$ - 2; however, the prototype will have a linear phase if $D_1 = n$ - 1. In this case, the input-output delay introduced by the analysis and synthesis banks is given by $D = F$ - 1 + D_0F (Section 4.5.1).

For the PR prototype filter, again it will be necessary to use its energy in the rejection range according to the objective function shown in equation (4.29), meeting the 2R restrictions imposed in equation (4.36). This type of filter performs well, in terms of implementation complexity [**118**]. We will use it in our second proposed work; BSS using NMD_UEPS.

4.7 Adaptive Scheme without Decimation

This scheme uses an analysis bank followed by adaptive filters with non-zero coefficients. This adaptive scheme is shown in Fig. 4.10 [**27**] and is also used in our design. This scheme is capable of modelling an FIR system due to the length of the analysis filters that is greater than the number of adaptive coefficients. However, it is shown in [**121**] that with the correct number of the adaptive filters taps $S_r(z^L)$ and the analysis filter bank $P_r(z)$,the scheme shown in Fig. 4.10 will be capable to model any FIR structure with some delay.

Considering the analysis polyphase bank representation of the scheme in Fig. 4.10, the polyphase matrix of dimension $L \times L$ is defined as [**27**]

$$P_m(z) = \begin{bmatrix} P_{0,0}(z) & P_{0,1}(z) & \Lambda & P_{0,L-1}(z) \\ P_{1,0}(z) & P_{1,1}(z) & \Lambda & P_{1,L-1}(z) \\ \text{M} & \text{M} & \text{O} & \text{M} \\ P_{L-1,0}(z) & P_{L-1,1}(z) & \Lambda & P_{L-1,L-1}(z) \end{bmatrix} \tag{4.38}$$

where $P_{r,c}(z)$ are polyphase components of type I of the r^{th} analysis filter $P_r(z) = \sum_{k=0}^{K_p-1} p_r(k) z^{-k}$, given by

$$P_{r,c}(z) = \sum_{k=0}^{\lfloor (K_p-1)/L \rfloor - 1} p_r(kL + c) z^{-k} \tag{4.39}$$

Therefore, the system function used in Fig. 4.10 can be expressed as

$$P(z) = \left[S_0(z^L) S_1(z^L) \Lambda \; S_{L-1}(z^L) \right] P_m(z^L) \begin{bmatrix} 1 \\ z^{-1} \\ \text{M} \\ z^{-(L-1)} \end{bmatrix} \tag{4.40}$$

The taps of the filters of non-zero coefficients $S_r(z^L)$ are changed to give the equivalent FIR scheme which will be called $U(z)$ [**27**]. The decomposition of the type I polyphase transfer function of the unknown system is given by

$$U(z) = \left[U_0(z^L)U_1(z^L)\Lambda\ U_{L-1}(z^L)\right]\begin{bmatrix} 1 \\ z^{-1} \\ \mathrm{M} \\ z^{-(L-1)} \end{bmatrix} \tag{4.41}$$

From equations (4.40) and (4.41), it can be noticed that the scheme accurately models an unknown FIR system when

$$\left[S_0(z^L)S_1(z^L)\Lambda\ S_{L-1}(z^L)\right]\mathrm{P}_m(z^L) = \left[U_0(z^L)U_1(z^L)\Lambda\ U_{L-1}(z^L)\right]. \tag{4.42}$$

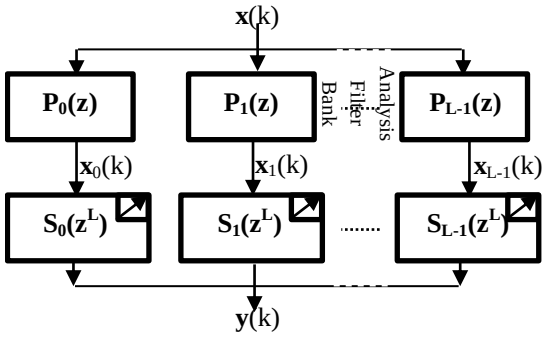


Fig. 4.10: Adaptive scheme employing analysis filters bank and non-zero coefficients filters [**27**].

Observing equation (4.42), we can see that the equality cannot be achieved as the length of the adaptive filters of non-zero coefficients is $L\kappa$ and the length of the analysis filters is K_p, while the product $S_r(z^L)P_{r,c}(z^L)$ has length $L\lceil K_p/L\rceil + L\kappa - 1$, which is greater than the number of coefficients $L\kappa$ that was adapted. However, if [**27**]

$$\left[S_0(z^L)S_1(z^L)\Lambda\ S_{L-1}(z^L)\right] = \left[U_0(z^L)U_1(z^L)\Lambda\ U_{L-1}(z^L)\right]Q_m(z^L) \tag{4.43}$$

such that $\mathbf{Q}_m(z^L)\mathbf{P}_m(z^L) = z^{-d}\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of dimension $L \times L$, the system function in Fig. 4.10 will have

$$P(z) = U(z)z^{-d}. \tag{4.44}$$

The matrices $\mathbf{P}_m(z)$ and $\mathbf{Q}_m(z)$ that satisfy the above conditions are, respectively, the polyphase matrix of the analysis and synthesis filter bank with perfect reconstruction. The synthesis polyphase bank matrix is defined as

$$\mathbf{Q}_m(z) = \begin{bmatrix} Q_{0,0}(z) & Q_{1,0}(z) & \Lambda & Q_{L-1,0}(z) \\ Q_{0,1}(z) & Q_{1,1}(z) & \Lambda & Q_{L-1,1}(z) \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ Q_{0,L-1}(z) & Q_{1,L-1}(z) & \Lambda & Q_{L-1,L-1}(z) \end{bmatrix}. \tag{4.45}$$

where $Q_{r,c}(z)$ are polyphase components of type II and the r^{th} synthesis filter $Q_r(z) = \sum_{k=0}^{K_q-1} q_r(k)z^{-k}$ is given by [**27**]

$$\mathbf{Q}_{r,c}(z) = \sum_{k=0}^{\lfloor (K_q-1)/L \rfloor} q_r\left(kL - c + L - 1\right)z^{-k}. \tag{4.46}$$

Then, using an analysis filter bank, which allows perfect reconstruction and adaptive filters of non-zero coefficients with sufficient order to satisfy equation (4.43), the scheme of Fig. 4.10, redrawn in Fig. 4.11, implements exactly the FIR system with the transfer function given in equation (4.44). However, it should be emphasised that the delay introduced by the filter bank must be considered in the algorithm for adapting the filters coefficients.

For analysis and synthesis filters of a linear phase of length $K_p = K_q = K_{pr}$, the delay is given by $d = K_{pr} - L$ [**122**]. The number of non-zero coefficients κ of the adaptive filters $S_r(z^L)$ must be at least:

$$\kappa = \lfloor (K_{un} + K_{pr})/L \rfloor - 1, \tag{4.47}$$

where K_{un} is the length of the unknown system.

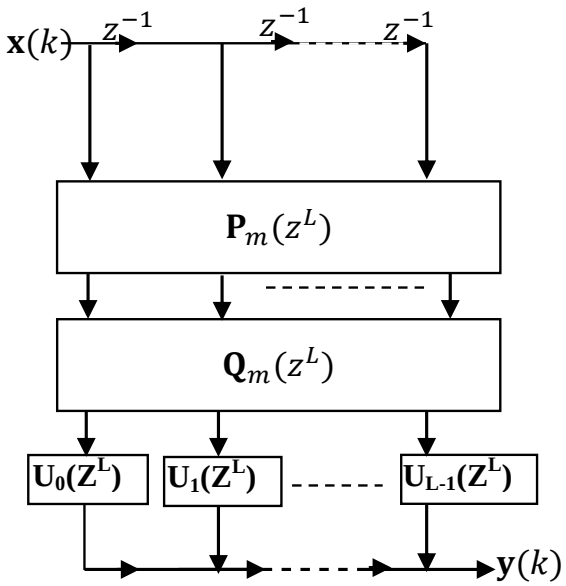


Fig. 4.11 Scheme for implementing FIR system with transfer function U(z)$z^{-\mathrm{d}}$.

4.8 Conclusions:

We conclude that an appropriate choice of the number of taps of the non-zero adaptive filters and the analysis filter bank can implement any FIR system. Furthermore, using polyphase representation for the analysis bank will give high computation efficiency in implementing both decimators and interpolators that are used in the filter banks. The ployphase structure of the analysis and synthesis filters operates at a rate *F* times smaller than the input signal *x*(*k*). The perfect reconstruction filter bank is the best choice, among other types discussed in this chapter, as its aliasing, amplitude distortion, and phase distortion can be avoided in the design.

Chapter 5
Maximally Decimated Adaptive Filtering with Multiple bands

5.1 Introduction

This chapter presents the first contribution of this research. The aim of this work is to remove the overlapping between neighbouring bands, improve the convergence rate and decrease the computational cost of the adaptive algorithms. The proposed work is a scheme that can be used in applications where a high number of taps are needed. In this work we extend the work of a MD_UEPS that uses a limited number of bands in [**26**] to an arbitrary number of bands. Our method is able to cancel the aliasing between neighbouring bands. Thus, an LMS algorithm with a modified step rate is formulated. This algorithm operates at a very small decimation and decreases the errors between bands. An analytical explanation is presented that elaborates on the evaluation of the MSE of the stationary state and speed of convergence. The proposed scheme is applied to system identification. An important enhancement is obtained in robustness, improving the speed of convergence and lowering the stationary error. The experiments showed the high performance of the proposed method against noisy inputs.

5.2 MD_UEPS with multiple bands

This section discusses in detail, the design procedure and formulation of the proposed method. The construction of the filters of UEPS with perfect reconstruction is also explored. The UEPS with multiple bands is derived from the scheme shown in Fig. 4.10, but by employing an UEPS analysis filter bank. Fig. 5.1 shows such a scheme, where $e(k)$ is the error utilised in the adaptation algorithm, $r(k)$ is the required signal, $P_i(z)$ are the UEPS analysis filters of L-bands, $x(k)$ is the input signal, and $S_i(z^{F_i})$ are the adaptive filters with non-zero coefficients.

The enhancement of a multiple bands scheme with a maximal decimation and of the corresponding adaptive approach will be accomplished for filter banks that use octave frequency decomposition. The low-pass features that usually exist in speech signals are behind this choice [**7**].

The perfect reconstruction analysis filter bank of L-bands is obtained as given in Section 4.4 (see equation (4.4)). These filters have orders

$$K_{P_0} = \sum_{c=0}^{L-2} 2^c K_{P^{0,c}}, \qquad K_{P_i} = \sum_{c=0}^{L-i-2} 2^c K_{P^{0,c}} + 2^{L-i-1} K_{P^{1,c}} \tag{5.1}$$

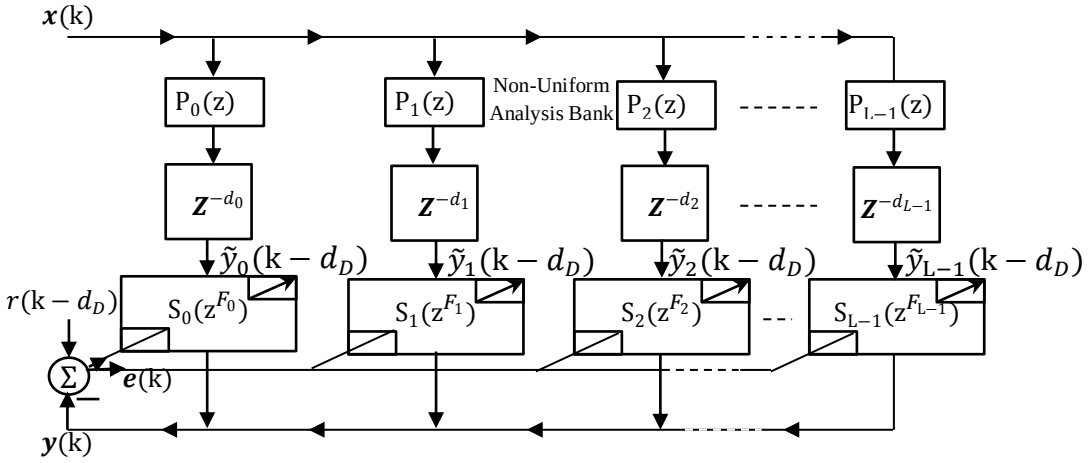


Fig. 5.1 Adaptive UEPS without decimation.

where $K_P^{0,c}$ are orders of $P^{0,c}(z)$ and $K_P^{1,c}$ are orders of $P^{1,c}(z)$. Fig. 5.1 shows the adaptive UEPS without decimation. The factors $\mathcal{F}_i$ that appear in the filters $S_i(z^{F_i})$ are defined in equation (4.6). The delays, d_i, that appear with z^{d_i} are necessary to compensate the introduction of different delays in the bands due to the differences in the length of the filters that compose the analysis bank [**7**] and are given by

$$d_i = K_{P_0} - K_{P_i} \quad . \tag{5.2}$$

The input-output total delay introduced by the scheme is

$$d_D = K_{P_0} \quad . \tag{5.3}$$

Our new maximally decimated structure is constructed using the following extension and assumptions:

Extension: add filters of UEPS with PR analysis $P_i(z)$ and synthesis filters $Q_i(z)$ to Fig. 5.1 after each of the sub-adaptive filter as explained in Fig. 5.2 that shows the i^{th} band of the resulting scheme. This extension allows, as discussed below, the filters to operate at a lower sampling rate.

Assumptions:

Assumption 1: To obtain a scheme with less complexity, we consider the analysis filters that are sufficiently selective to assume that there is spectrum interference only between the frequency responses of neighbouring bands. Non adjacent analysis filters, $P_i(z)$ and $P_j(z)$ for $|i - j| > 1$, have frequency responses that do not overlap;

Assumption 2: Assuming that $\frac{F_{i-1}}{F_i}$ is an integer[6];

Assumption 3: The coefficients of the filters $S_i(z^{F_i})$ vary slowly.

It can be seen that $P_{r,c}(z) = P_r(z)P_c(z)$ are the filters' coefficients $S_i(z^{F_i})$ shifted forward by F_i. After applying the extension and the above assumptions, we will have a simplified scheme of the i^{th} band that is shown in Fig. 5.3 . The sampling rate of the adaptive filters in the new scheme is F_i and F_{i+1} times less than the rate of the input signal. Appendix A.1 gives a detailed explanation of applying the extension and the assumptions in Fig. 5.2 to get Fig. 5.3 .

[6] For instance, $F_{i-1} = 2F_i$ for dyadic wavelets (special case of filter banks with UEPS) [106].

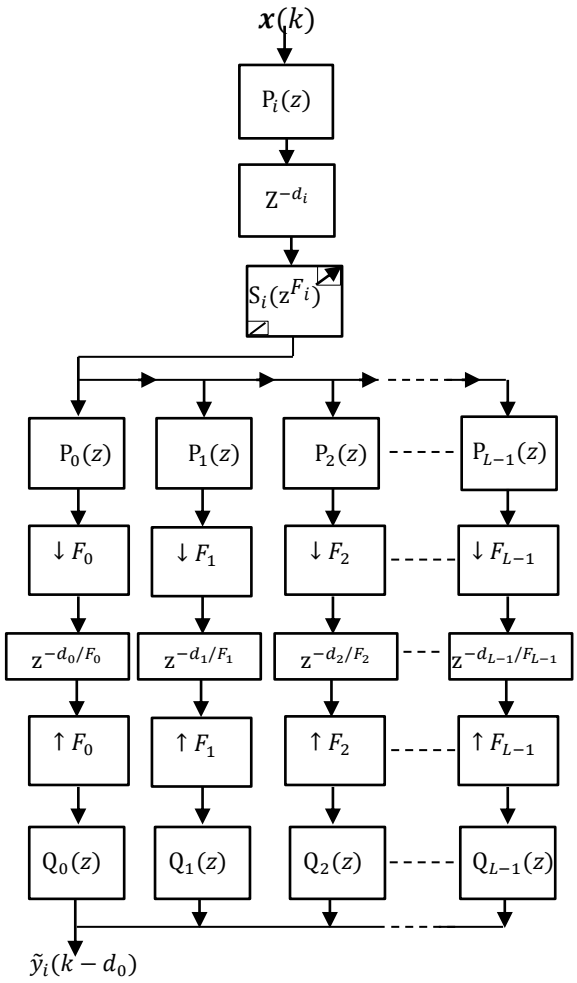


Fig. 5.2 Incorporating maximally decimated filter bank of the i^{th} band.

The scheme can be further simplified by noting that $P_{r,c}(z) = P_{c,r}(z)$ and combining the signals in adjacent bands. Fig. 5.4 illustrates the scheme for $L = 4$ bands, where $\eta(k)$ is a noise present in the desired signal $r(k)$. This figure illustrates the errors of each band $e_i(l)$, which will be used in Section 5.45.4 for derivation of the adaptation algorithm.

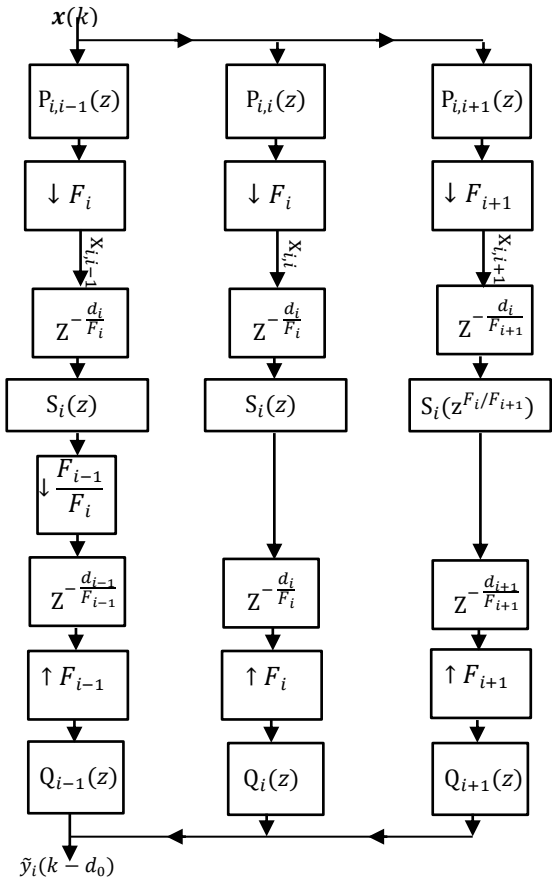


Fig. 5.3 i^{th} band adaptive filters working at lower rates.

5.3 Taps Selection

As mentioned in *Assumption*-1, Section 5.2, regarding the optimum design for an analysis filter that assumes that there is an overlapping spectrum between neighbouring bands, exclusively. The parameters of the maximally decimated scheme in Fig. 5.3 are similar to the parameters of the non-zero coefficients scheme in Fig. 5.1 . From this assumption, an equation will be formulated for the best selected taps of the MD_UEPS, assuming the case of modelling a random FIR system.

The adaptive filters $S_i(z^{F_i})$ for each band of Fig. 5.1 are described with regard to the shifted form of their $\ell_i = F_0/F_i$ polyphase elements [**110**], as illustrated in Fig. 5.5, leading to ℓ_i filters with F_0 components. Defining $S_i(z)$ as a vector that consists of ℓ_i polyphase elements of the i^{th} filter,

$$S_i(z) = \left[S_{i,0}(z) S_{i,1}(z) \Lambda \; S_{i,\lambda_{i-1}}(z)\right]^T, \tag{5.4}$$

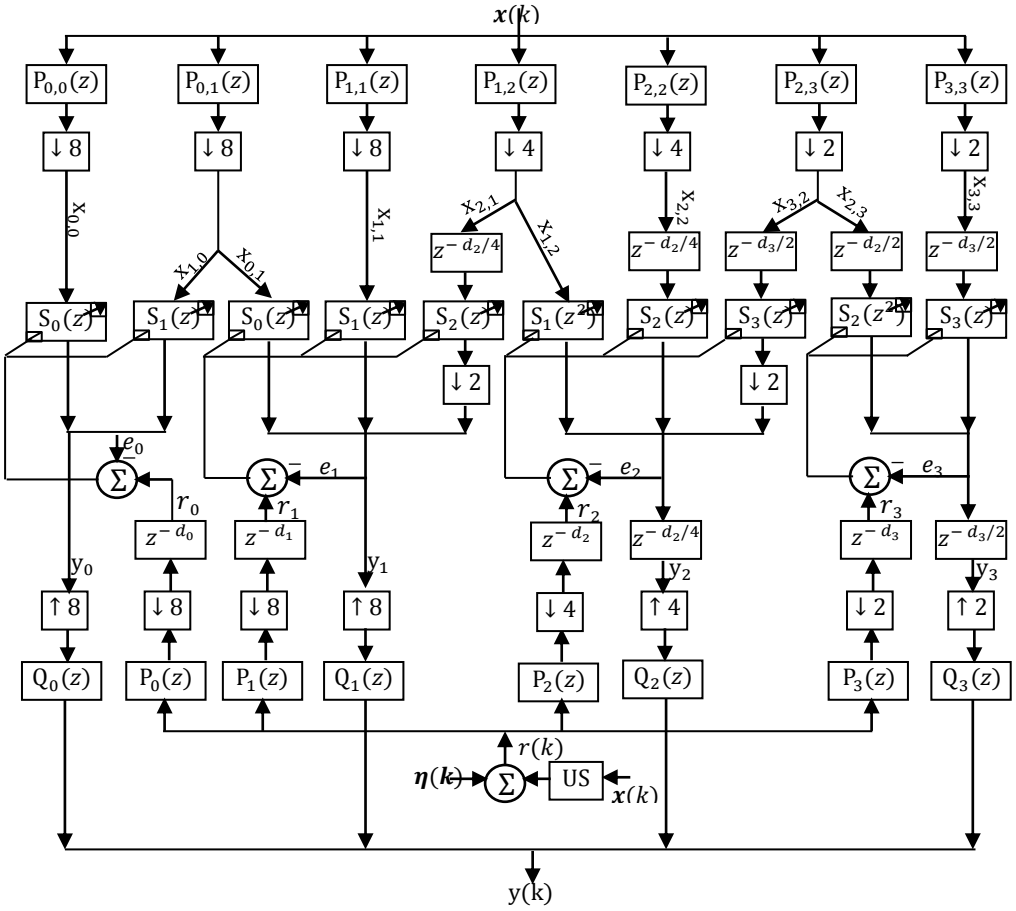


Fig. 5.4 Simplified scheme with L = 4 bands applied in the system identification.

and $\tilde{S}(z)$ is expressed in terms of the polyphase vectors of L filters $S_i(z)$, given by

$$\widetilde{S}(z) = \left[\widetilde{S}_0(z) \widetilde{S}_1(z) \Lambda \; \widetilde{S}_{F_0-1}(z)\right]^T == \left[S_0^T(z) S_1^T(z) \Lambda \; S_{L-1}^T(z)\right]^T \tag{5.5}$$

The multiband scheme of Fig. 5.1 can be amended to include $\mathcal{F}_0$ elements of $\tilde{\mathbf{S}}(z)$, as illustrated in Fig. 5.6. These elements are linked to the polyphase elements of filters the $S_i(z)$ by the following relations: $\tilde{S}_0(z) = S_{0,0}(z)$ and $\tilde{S}_{\ell_i+r}(z) = S_{i,r}(z)$ for i = 1, · · ·, L - 1 and r = 0, · · ·, ℓ_i - 1. For example, considering the UEPS with multiple bands of L = 4 (shown in Fig. 5.4), we get $\tilde{S}_0(z) = S_{0,0}(z)$, $\tilde{S}_1(z) = S_{1,0}(z)$, $\tilde{S}_2(z) =$

$S_{2,0}(z)$, $\tilde{S}_3(z) = S_{2,1}(z)$, $\tilde{S}_4(z) = S_{3,0}(z)$, $\tilde{S}_5(z) = S_{3,1}(z)$, $\tilde{S}_6(z) = S_{3,2}(z)$, and $\tilde{S}_7(z) = S_{3,3}(z)$. Appendix A.2 gives further explanation to these relations.

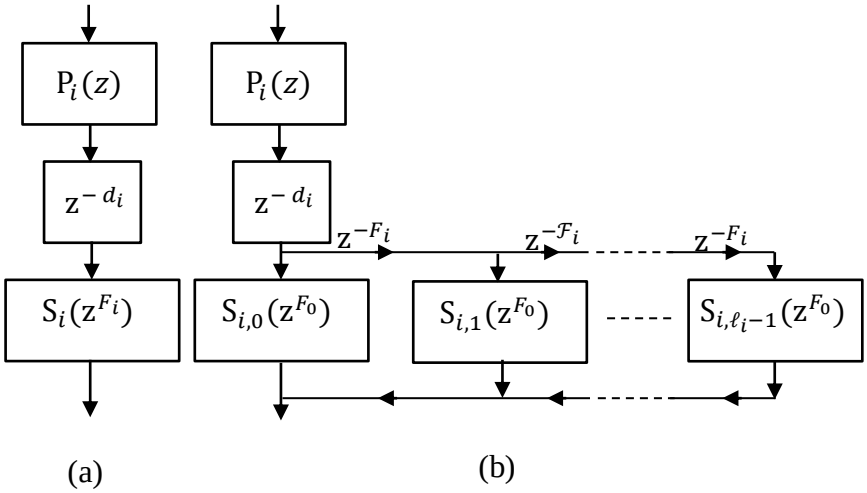


Fig. 5.5: (a) i^{th} band scheme of Fig. 5.1 (b) i^{th} band with the extended representation of $S_i(z^{F_i})$.

In Fig. 5.6, $\mathbf{P}_m(z)$ is the matrix of dimension F_0xF_0 that has the type-I polyphase elements of the analysis filters, written as

$$P_m(z) = \left[P_0^T(z) P_1^T(z) \Lambda \; P_{L-1}^T(z)\right]^T \tag{5.6}$$

where $\mathbf{P}_i(z)$ is the matrix ℓ_i x F_0 with the r^{th} row (r = 0, · · ·, ℓ_i - 1) formed by F_0 polyphase elements of $z^{-(d_i + rF_i)} P_i(z)$ (see Fig. 5.5). Looking at Fig. 5.6, the system function applied to the UEPS in Fig. 5.1 can be written with regard to the L_0 polyphase elements as :

$$T(z) = \tilde{S}^T\left(z^{F_0}\right) P_m\left(z^{F_0}\right)\left[1 \quad z^{-1} \quad \Lambda \quad z^{-(F_0-1)}\right]^T. \tag{5.7}$$

To identify the unknown system, taps of $S_i(z^{F_i})$ can be adjusted to model an FIR scheme. The system function of the unknown system is denoted by $U(z)$ and written as

$$U(z) = \left[U_0\left(z^{F_0}\right) \; U_1\left(z^{F_0}\right) \Lambda \; U_{L_0-1}\left(z^{F_0}\right)\right]\left[1 \quad z^{-1} \quad \Lambda \quad z^{-(F_0-1)}\right]^T. \tag{5.8}$$

From equations (5.7) and (5.8), the multiband scheme accurately models the FIR filter U(z) at

$$\widetilde{S}^{T}(z)P_m(z) = \begin{bmatrix} U_0(z) & U_1(z) & \Lambda & U_{F_0-1}(z) \end{bmatrix} \tag{5.9}$$

Post-multiplying both sides of equation (5.9) by the matrix $\mathbf{Q}_m(z)$ yields:

$$P_m(z)Q_m(z) = z^{-d_m} I, \tag{5.10}$$

where d_m is a positive integer and **I** is the identity matrix of dimension $F_0 \times F_0$. The following equation connects the non-zero filters and the unknown system parameters:

$$\widetilde{S}^{T}(z)z^{-d_m} = \begin{bmatrix} U_0(z) & U_1(z) & \Lambda & U_{F_0-1}(z) \end{bmatrix} Q_m(z). \tag{5.11}$$

The matrix $Q_m(z)$ that satisfies equation (5.10) corresponds to the synthesis polyphase filters matrix which leads to a system with perfect reconstruction [**111**]. Fig. 5.5(a) describes the i^{th} band of an analysis/synthesis system of UEPS with PR and Fig. 5.5(b) describes the extension with ℓ_i elements. The matrix $\mathbf{Q}_m(z)$ is of dimension $F_0 \times F_0$ containing the elements of the type-II polyphase of the expanded synthesis filters, given by

$$Q_m(z) = \begin{bmatrix} Q_0(z) & Q_1(z) & \Lambda & Q_{L-1}(z) \end{bmatrix}, \tag{5.12}$$

where $\mathbf{Q}_i(z)$ is the $F_0 \times \ell_i$ matrix with the r^{th} column (r = 0, · · ·, ℓ_i - 1) constructed by the F_0 type II polyphase elements of $z^{-(\ell_i - r - 1)F_i} Q_i(z)$.

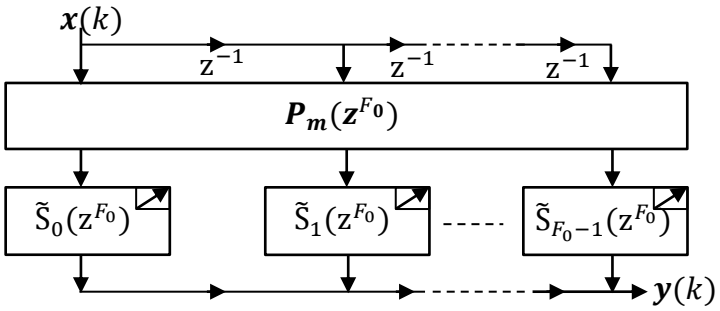


Fig. 5.6: Scheme of Fig. 5.1 using polyphase system.

The best selected taps of the i^{th} filter $\check{S}_i(z)$, assuming the existence of overlapping spectrum exclusively between neighbouring bands, are given by

$$\check{S}_i(z) = \sum_{r=0}^{\lambda_i - 1} z^{-r} S_{i,r}(z^{\lambda_i}) \tag{5.13}$$

where filters $S_{i,r}(z)$ are related to $\check{S}_i(z)$ through equations (5.4) and (5.5). According to equations (5.11) and (5.13) for an UEPS of L-band with synthesis filters that have order K_{Q_i}, we can write

$$K_{S_i} = \left\lceil \frac{K_U + K_{Q_i}}{F_i} \right\rceil + 1 \tag{5.14}$$

where K_{Si} is the minimum number of taps of the $S_i(z)$ filters and K_U is the required system order. Then, using a filter bank of UEPS that allows perfect reconstruction to be achieved and using adaptive filters of non-zero coefficients with orders that are sufficient to satisfy equation (5.14), the scheme in Fig. 5.3 can implement exactly any FIR system. However, it should be emphasised that the delay introduced by the filter bank should be considered in the adaptation algorithm. For more clarification, an example is given in Appendix A.3.

5.4 Adaptive Algorithm

In this section, we will derive a method based on the LMS algorithm to update the coefficients of the filters for the proposed MD_UEPS. The proposed adaptive algorithm works at a reduced sampling rate between the bands. The algorithm uses a novel adaptation step. This adaptation is normalised by the signals' powers at each band. We define $x_{r,c}(n)$ as the decimated signal at the output of the filters $P_{r,c}(z)$ (see Fig. 5.3) at iteration n, and $\mathbf{x}_{r,c}(n)$ has the latest samples K_{S_r} of $x_{r,c}(n)$ that was decimated by a factor of $F_{r,c} = \min(F_r, F_c)$ in relation to input signal $x(k)$, and $s_r(n)$ consists of the coefficients

of the filter $S_r(z)$. The objective function is considered as the sum of the mean squared errors[7] of each band, that is

$$\Im(n) = \sum_{i=0}^{L-1} \frac{1}{\lambda_i} \sum_{n=0}^{\lambda_i - 1} e_i^2(n_i') \tag{5.15}$$

where $n_i' = \ell_i n + k$ and $k = 0, \cdots, \ell_i - 1$. The update equation of the gradient descent type [**14**] for the coefficients of the i^{th} filter:

$$s_i(n+1) = s_i(n) - \frac{\gamma}{2} \frac{\partial \Im(n)}{\partial s_i}, \tag{5.16}$$

where γ is the learning rate. The error of each band is shown in Fig. 5.7, and given by

$$\begin{aligned} e_i(n_i') = r_i(n_i' - d_i) - x_{i,i}^T\left(n_i' - \frac{d_i}{F_i}\right)s_i(n) - x_{i-1,i}^T\left(n_i' - \frac{d_{i-1}}{F_i}\right)s_{i-1}(n) \\ - x_{i+1,i}^T\left(\frac{F_i}{F_i+1}n_i' - \frac{d_{i-1}}{F_i}\right)s_{i-1}(n) \end{aligned} \tag{5.17}$$

where $\vec{\mathbf{x}}_{i,i+1}$ is a vector that contains K_{S_i} samples of the signal $x_{i,i+1}(m)$ considering only $(F_i / F_{i+1})^{th}$ sample, due to the ratio (F_i / F_{i+1}) of the filter $S_i\left(z^{(F_i / F_{i+1})}\right)$ in Fig. 5.3 .

Substituting equation (5.17) in equation (5.15), we get

$$\begin{aligned} \frac{\partial \Im(n)}{\partial s_i} = \frac{2}{\lambda_i} \sum_{n=0}^{\lambda_i - 1} e_i(n_i') \frac{\partial e_i(n_i')}{\partial s_i} + \frac{2}{\lambda_{i-1}} \sum_{n=0}^{\lambda_{i-1} - 1} e_{i-1}(n_{i-1}') \frac{\partial e_{i-1}(n_{i-1}')}{\partial s_i} \\ + \frac{2}{\lambda_{i+1}} \sum_{n=0}^{\lambda_{i+1} - 1} e_{i+1}(n_{i+1}') \frac{\partial e_{i+1}(n_{i+1}')}{\partial s_i} \end{aligned} \tag{5.18}$$

[7] The adaptive filters work at different sampling rates, so we consider the mean of the errors of the samples at the outputs of the respective filters of the bands that operate at higher rates.

where

$$\frac{\partial e_i(n_i^{'})}{\partial s_i} = -x_{i,i}\left(n_i^{'} - \frac{d_i}{F_i}\right),$$

$$\frac{\partial e_{i-1}(n_{i-1}^{'})}{\partial s_i} = -x_{i,i-1}\left(\frac{F_{i-1}}{F_i}\, n_{i-1}^{'} - \frac{d_i}{F_{i-1}}\right), \qquad (5.19)$$

$$\frac{\partial e_{i+1}(n_{i+1}^{'})}{\partial s_i} = -\hat{x}_{i,i+1}\left(n_{i+1}^{'} - \frac{d_i}{F_{i+1}}\right)..$$

Substituting equation (5.18) and ((5.19) in equation (5.16) and considering different adaptation steps for each term, we obtain

$$s_i(n+1) = s_i(n) + \frac{\lambda_{i,i}}{\lambda_i}\sum_{n=0}^{\lambda_i - 1} x_{i,i}\left(n_i^{'} - \frac{d_i}{F_i}\right)e_i\left(n_i^{'}\right)$$

$$+ \frac{\gamma_{i-1,i}}{\lambda_{i-1}}\sum_{n=0}^{\lambda_{i-1}-1} x_{i,i-1}\left(\frac{F_{i-1}}{F_i}\, n_{i-1}^{'} - \frac{d_i}{F_{i-1}}\right)e_{i-1}\left(n_{i-1}^{'}\right) \qquad (5.20)$$

$$+ \frac{\gamma_{i,i+1}}{\lambda_{i+1}}\sum_{n=0}^{\lambda_{i+1}-1} \hat{x}_{i,i+1}\left(n_{i+1}^{'} - \frac{d_i}{F_{i+1}}\right)e_{i+1}\left(n_{i+1}^{'}\right)$$

To enhance the robustness and consequently the speed of convergence if the colour input is considered, every step of equation (5.20) is done conversely relative to the total input signals' powers, i.e.

$$\gamma_{i-1,i} = \frac{\tilde{\gamma}}{P_{i-2,i-1} + P_{i-1,i-1} + P_{i-1,i}},$$

$$\gamma_{i,i} = \frac{\tilde{\gamma}}{P_{i-1,i} + P_{i,i} + P_{i,i+1}}, \qquad 5.21)$$

$$\gamma_{i,i+1} = \frac{\tilde{\gamma}}{P_{i,i+1} + P_{i+1,i+1} + P_{i+1,i+2}},$$

where $\tilde{\gamma}$ is a constant factor and $P_{\lambda,u}$ is the power of the signal $x_{\lambda,u}(n)$ that can be predicted recursively by

$$P_{\lambda,u}(n+1) = \alpha P_{\lambda,u}(n) + (1-\alpha)x^2_{\lambda,u}(n) \tag{5.22}$$

with $0 \leq \alpha < 1$. For this reason, every term that represents the error in the recursive adaptation equation of (5.20) is measured by a specific normalised step value, as the error that is produced by the lower decimated signals has greater measure (weight).

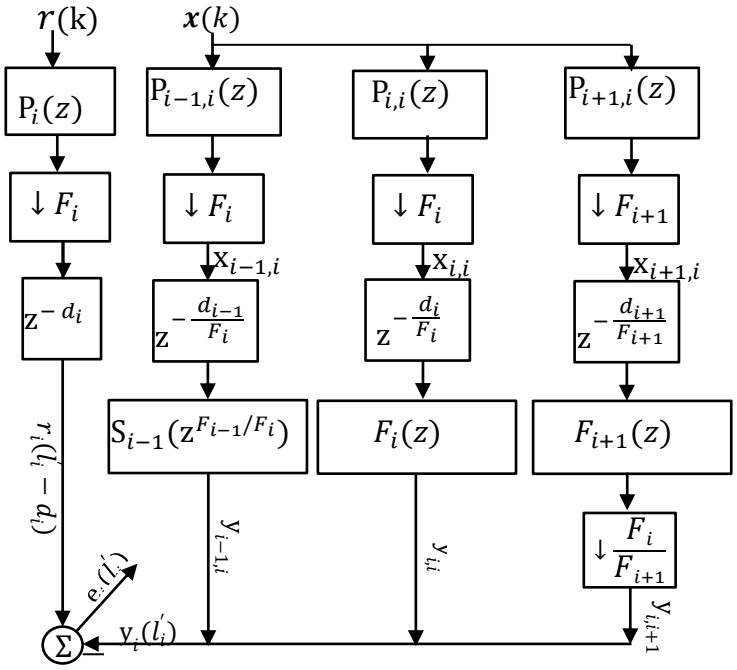


Fig. 5.7: Error of the i^{th} band.

5.5 Analysis of Convergence

To analyse the behaviour of the convergence of the adaptive algorithm in the previous section, we study the progress of the average error vectors of the coefficients, employing the independence theory and considering the stationary signals [**14**]. We can define the error vector of the coefficients of the i^{th} band as

$$\widetilde{s}_i(n) = s_i(n) - \text{ʕ}s_i \tag{5.23}$$

where $\check{s}_i$ is the corresponding vector of the best selected coefficients. Taking into account the assumption that the spectrum overlap of the analysis filters only happens between neighbouring bands (see Section 5.24.3) and that we modelled an FIR system, then the desired signals in each band $\mathfrak{d}_i$, after involving the coefficients $\check{s}_i$ can be formed as follows:

$$\begin{aligned} r_i\left(n_i^{'} - d_i\right) = x_{i,i}^T\left(n_i^{'} - \frac{d_i}{F_i}\right)\check{s}_i + \check{x}_{i-1,i}^T\left(n_i^{'} - \frac{d_{i-1}}{F_i}\right)\check{s}_{i-1} \\ + x_{i+1,i}^T\left(\frac{F_i}{F_{i+1}}n_i^{'} - \frac{d_{i+1}}{F_{i+1}}\right)\check{s}_{i+1} + \check{e}_i\left(n_i^{'}\right) \end{aligned} \tag{5.24}$$

where $\check{e}_i$ represents the modelling error[8] for the i^{th} band.

Let us consider vectors of dimension K_{S_r} x 1 containing the signals of the bands with their delays as follows:

$$\check{x}_{r,c}(n) = x_{r,c}\left(n - \frac{d_r}{F_{r,c}}\right) \tag{5.25}$$

$$\tilde{\check{x}}_{r,c}(n) = \check{x}_{r,c}\left(n - \frac{d_r}{F_{r,c}}\right) \tag{5.26}$$

where $F_{r,c} = \min(F_r, F_c)$. Therefore, substituting equations (5.17), (5.23) and (5.24) in equation (5.20), considering zero mean error of $\check{e}_i$, and taking the expected values of both sides of the resulting equation, we have

$$\begin{bmatrix} E\left[\tilde{s}_0(n+1)\right] \\ \text{M} \\ E\left[\tilde{s}_{L-1}(n+1)\right] \end{bmatrix} = \left((I - \Omega) \begin{bmatrix} E\left[\tilde{s}_0(n)\right] \\ \text{M} \\ E\left[\tilde{s}_{L-1}(n)\right] \end{bmatrix} \right) \tag{5.27}$$

where $\sum_{i=0}^{L-1} K_{S_i}$ represents the order of the unit matrix I. We can write

[8] error obtained with the optimal coefficients.

$$\Omega = \begin{bmatrix} \Theta 3_0 & \Theta 4_0 & \Theta 5_0 & 0 & \Lambda & \Lambda & \Lambda & 0 \\ \Theta 2_1 & \Theta 3_1 & \Theta 4_1 & \Theta 5_1 & 0 & O & \Lambda & M \\ \Theta 1_2 & \Theta 2_2 & \Theta 3_2 & \Theta 4_2 & \Theta 5_2 & 0 & O & M \\ 0 & O & O & O & O & O & O & M \\ M & O & O & O & O & O & O & 0 \\ M & O & 0 & \Theta 1_{L-3} & \Theta 2_{L-3} & \Theta 3_{L-3} & \Theta 4_{L-3} & \Theta 5_{L-3} \\ M & \Lambda & O & 0 & \Theta 1_{L-2} & \Theta 2_{L-2} & \Theta 3_{L-2} & \Theta 4_{L-2} \\ 0 & \Lambda & \Lambda & \Lambda & 0 & \Theta 1_{L-1} & \Theta 2_{L-1} & \Theta 3_{L-1} \end{bmatrix} \tag{5.28}$$

where

$$\begin{aligned} \Theta 1_i &= \gamma_{i-1,i} E\left[\tilde{X}_{i,i-1}(n'_{i-1})\tilde{\mathfrak{P}}^T_{i-2,i-1}(n'_{i-1})\right], \\ \Theta 2_i &= \gamma_{i-1,i} E\left[\tilde{x}_{i,i-1}(n'_{i-1})\tilde{x}^T_{i-1,i-1}(n'_{i-1})\right] + \gamma_{i,i} E\left[\tilde{x}_{i,i}(n'_i)\tilde{\mathfrak{P}}^T_{i-1,i}(n'_i)\right], \\ \Theta 3_i &= \gamma_{i-1,i} E\left[\tilde{x}_{i,i-1}(n'_{i-1})\tilde{x}^T_{i,i-1}(n'_{i-1})\right] + \gamma_{i,i} E\left[\tilde{x}_{i,i}(n'_i)\tilde{x}^T_{i,i}(n'_i)\right] + \gamma_{i,i+1} E\left[\tilde{\mathfrak{P}}_{i,i+1}(n'_{i+1})\tilde{\mathfrak{P}}^T_{i,i+1}(n'_{i+1})\right] \\ \Theta 4_i &= \gamma_{i,i} E\left[\tilde{x}_{i,i}(n'_i)\tilde{x}^T_{i+1,i}(n'_i)\right] + \gamma_{i,i+1} E\left[\tilde{\mathfrak{P}}^T_{i,i+1}(n'_{i+1})\tilde{x}^T_{i+1,i+1}(n'_{i+1})\right], \\ \Theta 5_i &= \gamma_{i,i+1} E\left[\tilde{\mathfrak{P}}^T_{i,i+1}(n'_{i+1})\tilde{x}^T_{i+2,i+1}(n'_{i+1})\right], \end{aligned} \tag{5.29}$$

where the dimensions of the matrices $\Theta 1_i$, $\Theta 2_i$, $\Theta 3_i$, $\Theta 4_i$, and $\Theta 5_i$ are $K_{S_i} \mathrm{x}\, K_{S_{i-2}}$, $K_{S_i} \mathrm{x}\, K_{S_{i-1}}$, $K_{S_i} \mathrm{x}\, K_{S_i}$, $K_{S_i} \mathrm{x}\, K_{S_{i+1}}$ and $K_{S_i} \mathrm{x}\, K_{S_{i+2}}$, respectively.

The above matrices are expressed as follows[9]

$$E\left[\tilde{X}_{r,c}(n'_c)\tilde{X}^T_{\lambda,u}(n'_u)\right] = \tilde{P}_{r,c} R_{xx} \tilde{P}^T_{\lambda,u}, \tag{5.30}$$

and

$$E\left[\tilde{X}_{r,c}(n'_c)\tilde{\mathfrak{P}}^T_{\lambda,u}(n'_u)\right] = \tilde{P}_{r,c} R_{xx} \tilde{\mathfrak{P}}^T_{\lambda,u}. \tag{5.31}$$

[9] The matrices are formed to include the autocorrelation function of the input signal and the coefficients of the analysis filters.

where $\mathbf{R}_{xx}$ represents the autocorrelation of the input signal of dimension $K_T \times K_T$, as $K_T = \max\left[\left(F_c K_{S_r} + K_{P_{r,c}} + 1 + d_r\right), \left(F_u K_{S_\lambda} + K_{P_{\lambda,u}} + 1 + d_\lambda\right)\right]$; $\widetilde{\mathbf{P}}_{r,c}$ is a matrix of dimension $K_{S_c} \times K_T$, with the first row of non-zero elements at position $d_r + 1$ and containing the coefficients of $P_{r,c}(z)$ with the following rows being formed by shifting the elements of the previous row elements to the right position of F_c, and $\widetilde{\mathbf{P}}_{r,c}$ is similar to $\widetilde{\mathrm{P}}_{r,c}$ but with shifting of F_r positions from one row to the next.

By observing equation (5.27), we can deduce that the convergence rate of in equation (5.20), is controlled through the ratio between the eigenvalues of Ω. This result tells us how to predict the behaviour of the algorithm for a specific analysis bank and can also be utilised if some statistical information is available about the input signal, in order to choose the correct number of bands with regard to the full-band algorithm.

In Section 5.7, it will be demonstrated that the eigenvalue ratio of matrix Ω is notably reduced for coloured input signals when the number of bands is increased, resulting in a better convergence rate than that obtained by the full-band algorithm.

5.6 Steady state MSE

The mean square error will be analysed in a steady state of our suggested algorithm of the UEPS with multiple bands, which considers only the error caused by the assumption that there is no spectral overlaps between non-adjacent bands of the analysis filters, during the derivation of the adaptive algorithm (Section 5.4). Other errors can be modelled by conventional LMS analysis. Analysing the general case (i.e., considering that there is overlapping spectrum between non-adjacent bands), the desired signal $r_i(n_i' - \mathrm{d}_i)$ in the i^{th} band can be written by including the optimal coefficients $\check{s}_i$ at n as

$$r_i\left(n_i^{'} - d_i\right) = \sum_{r=0}^{i-1} \hat{x}_{r,i}^T \left(n_i^{'} - \frac{d_r}{F_i} \right) \check{s}_r + \sum_{r=0}^{L-1} x_{r,i}^T \left(\frac{F_i}{F_r} n_i^{'} - \frac{d_r}{F_i} \right) \check{s}_r + \eta_i\left(n_i^{'}\right) \tag{5.32}$$

where $\eta_i(n)$ is the measurement noise of the i^{th} band.

Defining $r(l) = [r_0(\ell_0 n)\ r_1(\ell_1 n) \cdots r_{L-1}(\ell_{L-1} n)]^T$ for L bands in $\ell_i n$ time, this is written by:

$$r(n) = X(n)\check{s} + \eta(n) \tag{5.33}$$

where $\check{s} = [\check{s}_0^T\ \check{s}_1^T\ \cdots \check{s}_{L-1}^T]^T$ represents the vector that consists of all optimal adaptive filters' coefficients, $\boldsymbol{\eta}(n)$ is the vector formed by residual errors modelling of all L bands in $\ell_i n$ time, and X(n) is a matrix of dimension L x $\left(\sum_{i=0}^{L-1} K_{S_i}\right)$ with input signals of the bands whose i^{th} row:

$$[X(n)]_i = \left[\tilde{x}_{0,i}^T(\lambda_i n)\ \ \Lambda\ \ \tilde{x}_{i-1,i}^T(\lambda_i n)\ \ \tilde{x}_{i,i}^T(\lambda_i n)\ \ \tilde{x}_{i+1,i}^T(2\lambda_i n)\ \ \Lambda\ \ \tilde{x}_{L-1,i}^T(2^{L-1-i}\lambda_i n)\right] \tag{5.34}$$

Now taking into account only the overlapping spectrum between neighbouring bands, the vector $y(n)$ is expressed by the output signals of the L bands of the UEPS in time $\ell_i n$ (see Fig. 5.4) as follows:

$$y(n) = \tilde{X}(n)s(n) \tag{5.35}$$

as $\tilde{\mathrm{X}}(n)$ can be expressed in terms of the input vectors of the adaptive filters:

$$\left[\tilde{X}(n)\right]_k = \left[0_0^T\ \Lambda\ 0_{i-2}^T\ \ \tilde{x}_{i-1,i}^T(\lambda_i n)\ \ \tilde{x}_{i,i}^T(\lambda_i n)\ \ \tilde{x}_{i+1,i}^T(2\lambda_i n)\ \ 0_{i+2}^T\ \Lambda\ 0_{L-1}^T\right] \tag{5.36}$$

and $s(n) = [s_0^T(n) s_1^T(n) \cdots s_{L-1}^T(n)]^T$ contains coefficients with 0_r^T vector of dimension 1 x K_{S_r} that is formed by replacing the zeros signal vector at the input of the r^{th} adaptive filter of a non-adjacent band. Therefore, the errors vector, that is obtained from the assumption that states; there is no overlapping between the spectrum of the non-neighboured analysis filters in the extraction of the maximally decimated scheme, is given by:

$$e(n) = r(n) - y(n) = X(n)\check{s} - \tilde{X}(n)s(n) + \eta(n). \tag{5.37}$$

where $\boldsymbol{\eta}(n)$ is the vector containing the residual modelling errors of each band.

A filter bank is considered to be lossless when the error power at the output is equal to the aggregate of the error powers in each band. Therefore, the entire mean square error [**123**] is given by:

$$\varsigma(n) = \sum_{i=0}^{L-1} \frac{1}{\lambda_i} \sum_{n=0}^{\lambda_i - 1} E\left[e_i^2(n_i')\right] \approx \sum_{i=0}^{L-1} E\left[e_i^2(\lambda_i n)\right] = E\left[e(n)^T e(n)\right] \quad (5.38)$$

Substituting equations (5.33), (5.35) and (5.37) in (5.38) and considering $\eta((n)) = 0$ and $E[s(\infty)] \approx \check{s}$, yields

$$\varsigma_{UU} = \varsigma(\infty) \approx \check{S}^T \Psi \check{S} + \sigma_\eta^2 \quad (5.39)$$

where

$$\Psi = E\left\{ \left[X(n) - \tilde{X}(n)\right] \left[X(n) - \tilde{X}(n)\right]^T \right\} \quad (5.40)$$

and σ_η^2 represents the variance of the measurement noise, and

$$\left[X(n) - \tilde{X}(n)\right]_i = \begin{bmatrix} \tilde{x}_{0,i}^T(\lambda_i n) & \Lambda & \tilde{x}_{i-2,i}^T(\lambda_i n) \\ 0_{i-1}^T & 0_i^T & 0_{i+1}^T \quad \tilde{x}_{i+2,i}^T(2\lambda_i n) \quad \Lambda \quad \tilde{x}_{L-1,i}^T(2^{L-1-i}\lambda_i n) \end{bmatrix} \quad (5.41)$$

consists of the input elements of the i^{th} band of non-neighbouring bands. Thus, Ψ can be expressed using the correlation matrices of the non-neighboured bands' signals.

The adaptive parameters of selective analysis filters will converge roughly towards the optimal coefficients of the non-zero coefficients filters. Consequently, the elements of the vector $\check{s}$ will be almost identical to those of the filters $\check{S}_i$ given by equation (5.13). It can be seen from equation (5.39) that the steady state MSE of the UEPS with multiple bands will be generally, greater than σ_η^2, as the residual spectral overlap is not removed in the simplified scheme. On the other hand, the stopband attenuation of the analysis filters is behind the corresponding increment in the steady state MSE that is evaluated by the following simplified equation:

$$\varsigma_{UU} \approx \nu_{av}(\Psi) \left|\check{S}\right|^2 + \sigma_\eta^2 \quad (5.42)$$

where $\nu_{av}(\Psi)$ represents the average of the eigenvalues of the matrix Ψ. This simplified way of evaluating the excess mean square error is not dependent on prior information about the optimal coefficients; the only information needed concerns the corresponding vector $\check{s}$. The vector $\check{s}$ can be obtained by finding the power ratio between the input and

output signals. Equation (5.42) also represents the relationship between the input autocorrelation sequence and the parameters of the analysis filters.

5.7 Results

In this section, using computer simulations, we compare the performance of the proposed UEPS with multiple bands (Section 5.4) to the MD_EPS of [**42**], and the NMD_UEPS of [**43**].

5.7.1 *Experiment 1*

We consider an FIR system identification of order KU = 900. The unknown system coefficients are attained arbitrarily (uniformly-distributed white signal). The input signal was produced by passing a white Gaussian signal through a first order IIR filter with pole at z = 0. 95. A white noise with variance $\sigma_k^2 = 10^{-6}$ is provided to the required signal. Our proposed structure was implemented by employing perfect reconstruction filter banks, and a decomposition in octaves with L = 3, 4, and 5 (see Chapter 4, Section 4.2, and [**108**]).

Table 5.1 presents the decimation factors F_i, the factors $\ell_i = F_0/F_i$, the delays d_i, and the orders K_{S_i} and K_{P_i} of the adaptive filters $S_i(z)$ and the analysis filters $P_i(z)$, respectively, that are employed in the experiment with L = 4 bands. Fig. 5.8 describes the frequency responses of the corresponding analysis filters. Trial and error is used to reach the best convergence from the adaptation step, $\breve{\gamma} = 1/K_{S0}$, of equation (5.21).

The practical MSE of the proposed UEPS is evaluated for various bands to ascertain the number of bands that provides the best performance; shown in Fig. 5.9. Table 5.2 presents major and minor eigenvalues of the matrices Ω (equation (5.28) for the simulations in Fig. 5.9.

When the theoretical results are analysed, it is assumed that the coefficients will be updated at a sampling rate of $F_0 = 2^{L-1}$ times less than the rate of the input signal. Thus, the useful eigenvalues ratios are explained in Table 5.2 multiplied by F_0. Considering this perception, we can confirm that the practical results in Fig. 5.9 are consistent with the theoretical outcomes in Table 5.2, showing that the convergence rate enhances when the number of bands increases compared to the full-band least mean square algorithm (L = 1 in Fig. 5.9). It is proved experimentally that four bands are enough to decorrelate the coloured signal that is used at the input. The little increase that is expected in the convergence speed when adding an additional decomposition level (L=5) has not been verified in practice, as the aggregate number of the adaptive coefficients of the multiband scheme increases. The effect of this increase in the number of adaptive coefficients will not be noticeable in the modelling of very high order systems like rooms and auditoriums with echo.

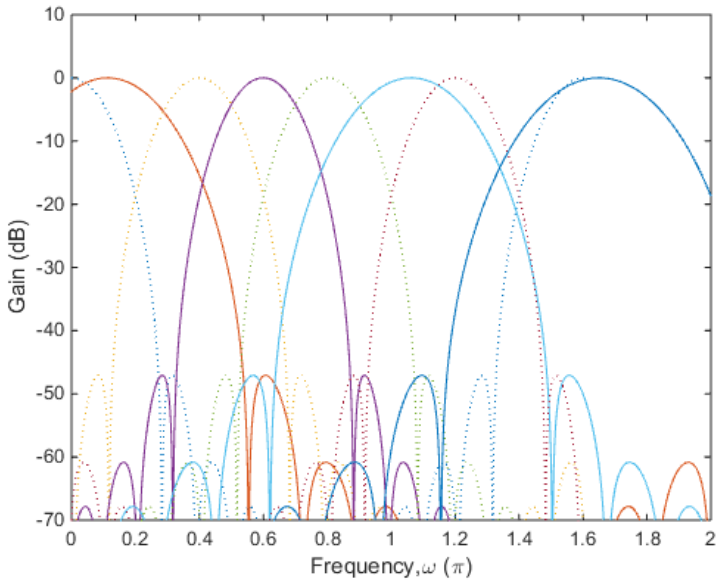


Fig. 5.8 Gain of analysis filters in Fig. 5.4 ($P_{i,i}(z)$: solid line, and $P_{i,i+1}(z)$: dashed line).

Table 5.1: Parameters of the scheme with UEPS for Experiment 1.

i	F_i	ℓ_i	K_{P_i}	d_i	K_{S_i}
0	8	1	332	0	149
1	4	2	332	0	149
2	4	2	155	163	256
3	2	4	59	267	480

Table 5.2: Ratio between the highest and lowest eigenvalue Ω for Experiment 1for the coloured input.

L	ν_{max}/ν_{min}
1	286. 40
3	37.15
4	8.04
5	2.99

Next, after proving that our proposed MD_UEPS with four bands (L=4) has a better convergence rate than the conventional full-band (L=1), we will validate our work by using the MSE evolution as a benchmark to compare with the current state-of-the-art. The comparison is shown in Fig. 5.10 for $L = 4$ bands; it compares the performance of the mean square error of our proposed method with the MD_EPS in [**42**] and the non-maximally (oversampled) decimated UEPS in [**43**]. It can be seen that the suggested

scheme with UEPS has the fastest convergence speed. To decrease the convergence problem of the oversampled scheme, we use analysis filters with wider bands. The use of UEPS over the EPS allows us to improve the convergence rate. The reason for the increase in the convergence rate of the UEPS over the EPS, is due to the use of finer decomposition at low frequencies of the input signal; this decreases the ratio of power spectral density of the bands' signals [**1**].

In this experiment, we improved the convergence obtained with the proposed UEPS with multiple bands using a filter bank in octaves with decomposition that was observed for coloured input with low-pass characteristics. On the other hand, increasing the number of decomposition levels for coloured input with high-pass characteristics would improve the convergence rate.

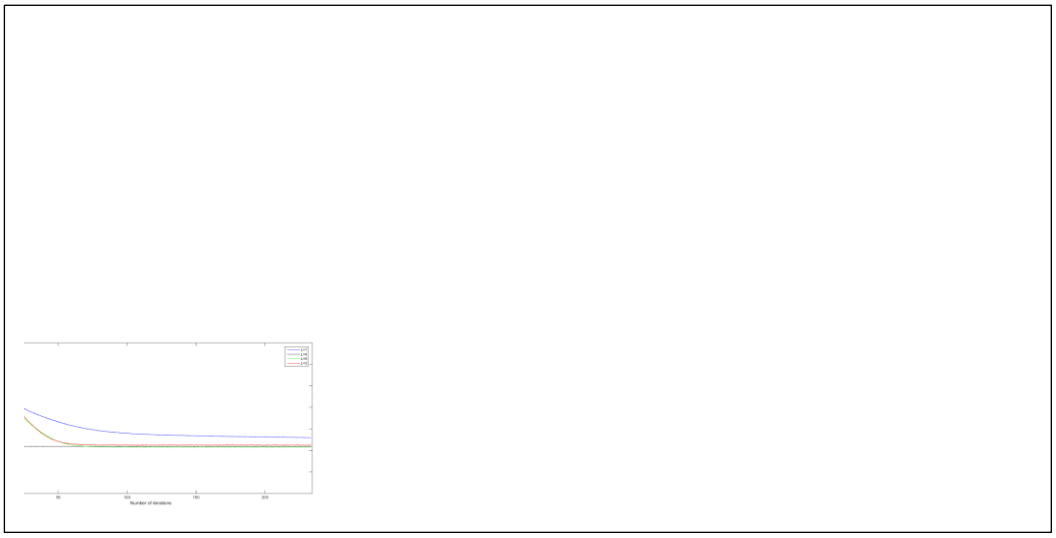


Fig. 5.9: Performance of the mean square error of the UEPS of Experiment 1.

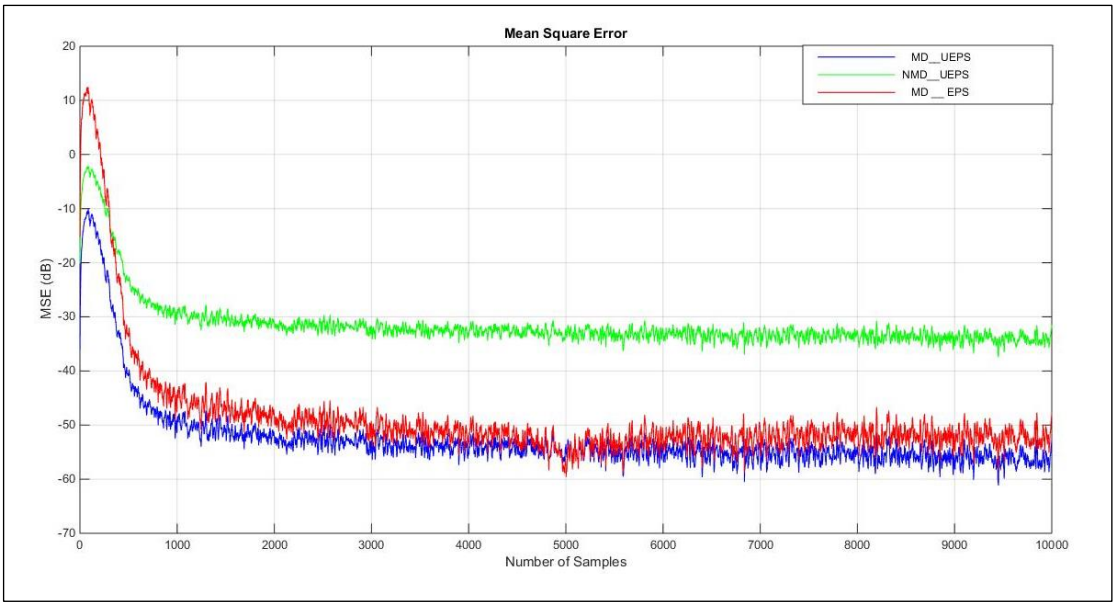


Fig. 5.10: Comparison of the mean square error of the proposed MD_UEPS to the MD_EPS of [**42**] and NMD_UEPS of [**43**], for L = 4 in Experiment 1.

5.7.2 *Experiment 2*

In this test, we used an FIR system identification of a similar order K_U = 900 to that in the first experiment. However, with the input signal was produced by applying Gaussian white noise over an IIR filter of a first order and pole at z = 0.98.

The evolution of the mean square error of the proposed UEPS with multiple bands with L = 4 and 5 bands is shown in Fig. 5.11. The eigenvalue spread is shown in Table 5.3, for the corresponding matrix Ω of equation (5.28), for the simulations of Fig. 5.11. It can be seen in Table 5.3 and Fig. 5.11 that the proposed scheme with L = 4 bands is not suitable for reducing the autocorrelation of the coloured input signal in this test. However, further increase in the number of bands to L = 5 will result in a slight increase in the convergence rate.

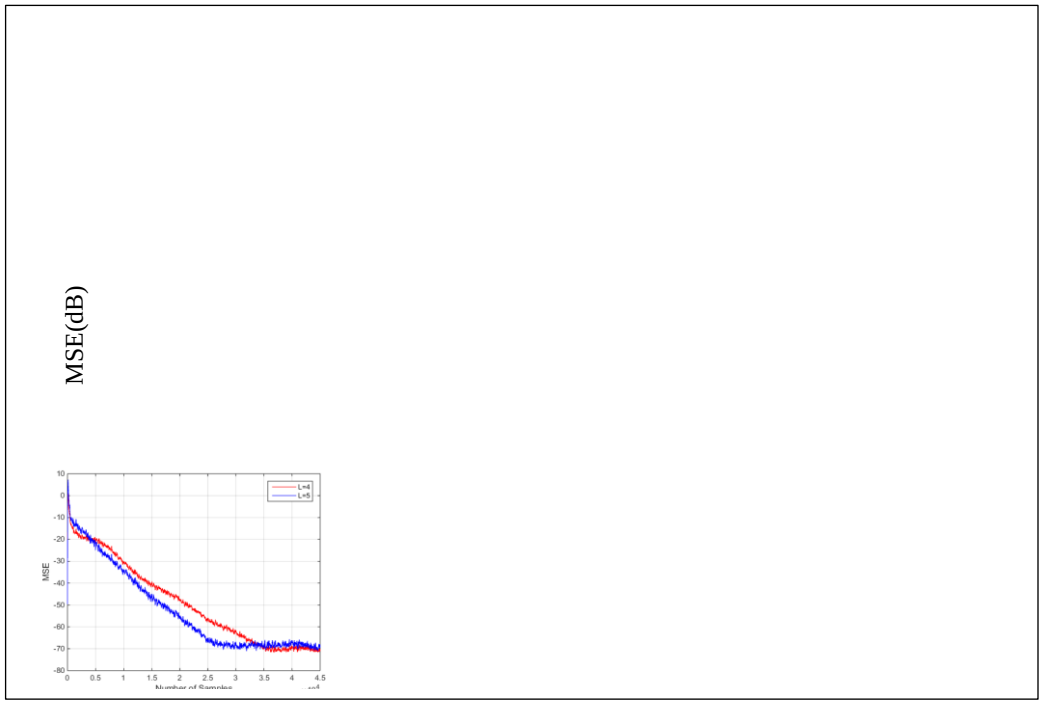


Fig. 5.11 : MSE with UEPS for Experiment 2.

Table 5.3: Eigenvalues of the matrix.

L	ν_{max}/ν_{min}
4	33.38
5	9.47

5.7.3 *Experiment 3*

The goal of this experiment is to evaluate the ability of our proposed UEPS to rapidly react to changes in the unknown system. At first, we studied the system identification of a lowpass FIR of order K_U = 299 with 0.4π bandwidth. The unknown system, after 15000 samples, was altered to a LPF of the same order, but with 0.2π

bandwidth. A coloured noise was generated at the input, as in the first experiment. A Gaussian white noise of $\sigma_{\eta}^2 = 10^{-5}$ and mean = 0 was added to the required signal. The scheme was carried out utilising the same perfect reconstruction filter bank of octave bands containing four bands from the first experiment. Fig. 5.12 shows the mean square error evolutions of our proposed MD_UEPS with multiple bands and the MD_EPS [**42**]. We can see from this figure that the proposed UEPS with multiple bands is reliable to keep abreast of the quick changes in the unknown system and has the same performance as the EPS.

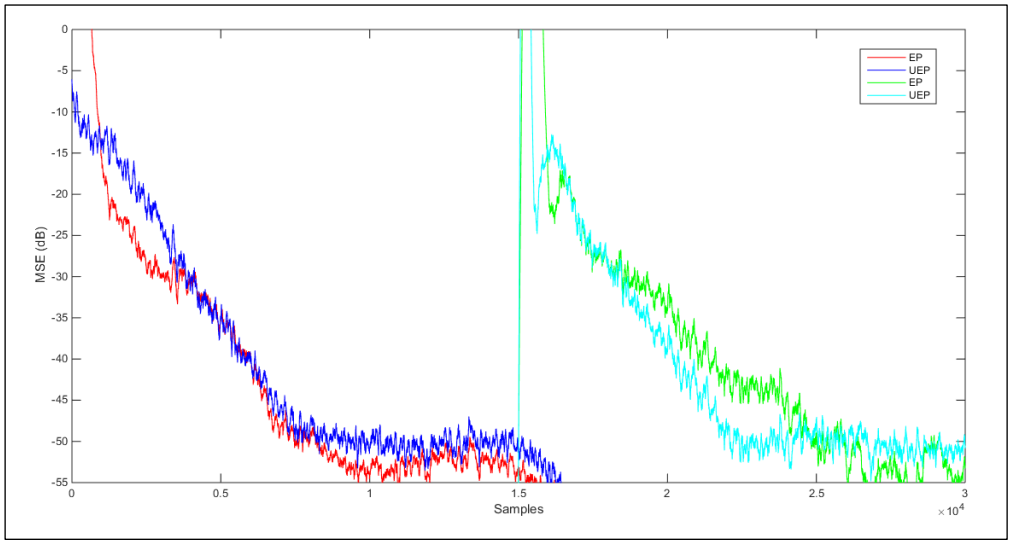


Fig. 5.12: MSE of the schemes with EPB [**42**] and the proposed UEPB for Experiment 3.

5.7.4 *Experiment 4*

In the previous experiments, we compared our proposed method with the full-band scheme. The aim of this experiment is to choose the proper prototype.

The mean square errors in the steady-state that is obtained experimentally and theoretically are compared in the system identification of order K_U =127 (with arbitrarily produced coefficients). The types of signals that are applied at the input were coloured

and white. In the coloured case, there is no measurement noise introduced to the required signal. The simulation of the proposed UEPS with multiple bands was performed using four bands, using tree structured filter banks with 3-level octave bands. To show that the desired steady-state MSE depends on the prototype filters order used to generate the filter banks, we choose the following orders of the prototype filters: K_{P_3}=15, 31, and 63. The three prototype filters with the results obtained on the steady-state mean square error (equation (5.42)) and the theoretical estimates (equation (5.39)) are shown in Table 5.4. The theoretical steady-states and the respective experimental mean square error evaluations of these prototype filters for UEPS with multiple bands are illustrated Fig. 5.13 and Fig. 5.14 for white and coloured inputs, respectively.

The frequency responses and their corresponding mean square errors results (see Table 5.4) for the three prototypes used in this experiment for coloured signal applied at the input are shown in Fig. 5.15.

We can see from Table 5.4, Fig. 5.13 and Fig. 5.14 that both the experimental and theoretical results are almost identical. The correct filter banks can be chosen with the help of equations (5.39) and (5.42) for a desired mean square error in a steady-state.

Table 5.4: Mean square errors in dB of theoretical and practical for Experiment 4.

Input	white noise			colour signal		
K_{P_3}	15	31	63	15	31	63
ζ_{UU} (equation (5.39))	-38.0	-60.0	-87.2	-37.1	-56.5	-81.3
ζ_{UU} (equation (5.42))	-36.8	-59.0	-86.5	-36.1	-56.2	-81.3
ζ_{exp}	-37.1	-61.0	-88.6	-32.1	-54.0	-80.2

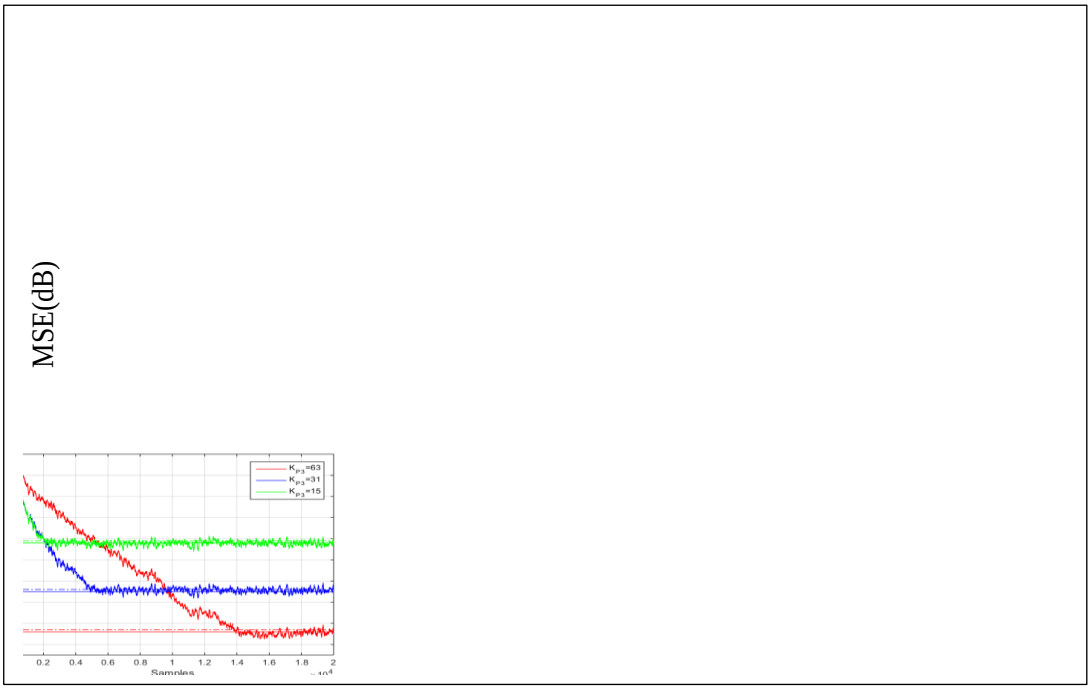


Fig. 5.13: MSE and theoretical values of the excess MSE for Experiment 4, considering a white input signal.(dashed lines for equation (5.37) and dotted lines for equation (5.40))

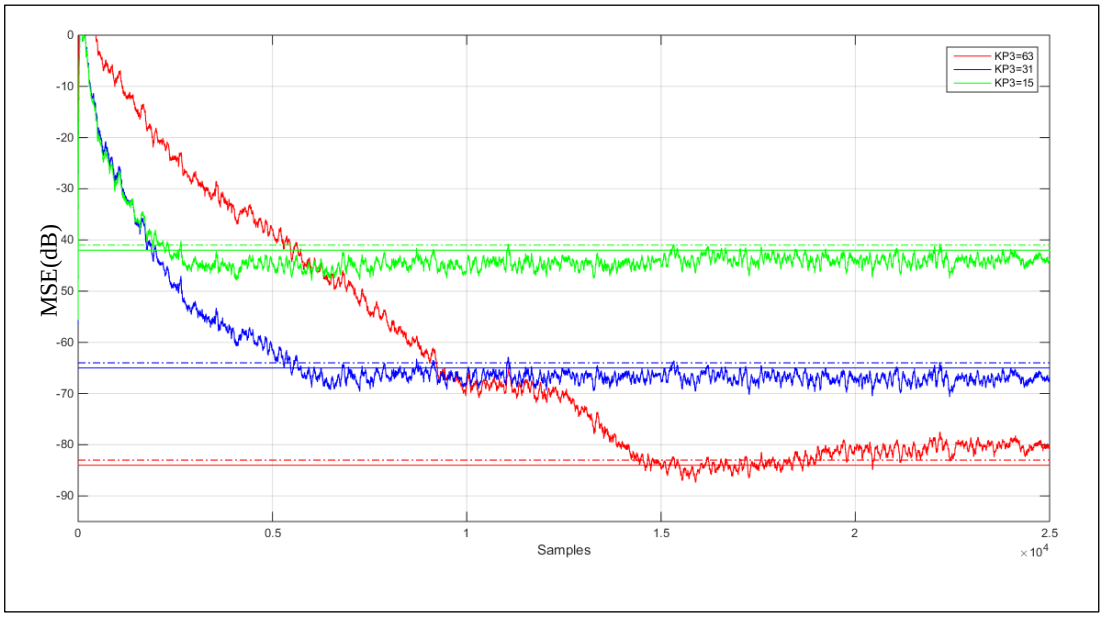


Fig. 5.14: MSE and theoretical values for the excess MSE in Experiment 4, considering a coloured input signal.

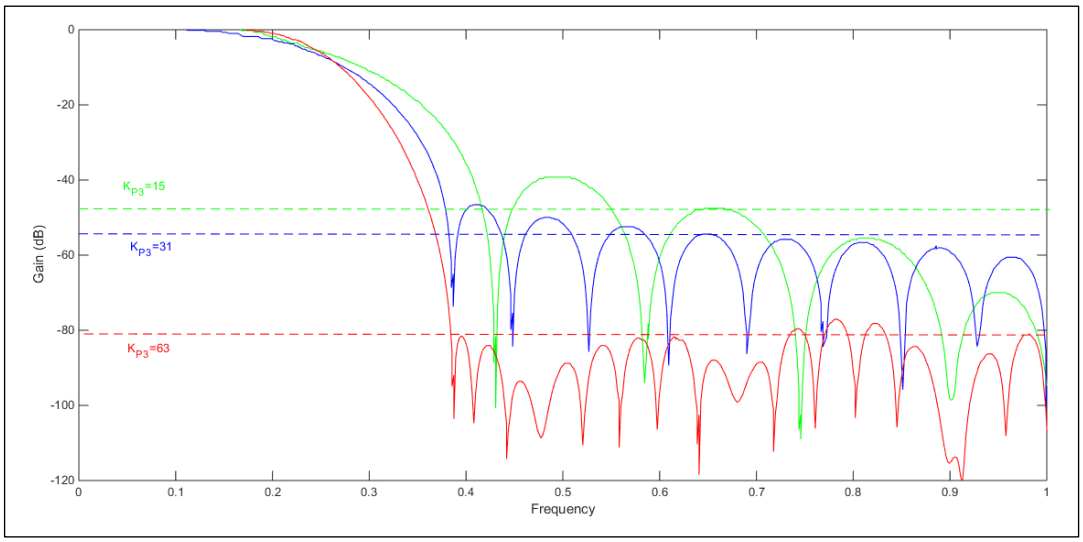


Fig. 5.15: Gain of prototype filter (solid line) and their MSEs experimental (dashed line), considering a coloured input signal.

5.8 Conclusions

This chapter discusses the decomposition of the input signal using a maximally decimated multiband scheme that uses a tree filter bank of UEPS. The derivation of this scheme is presented in detail. A normalised step-size least mean square algorithm is used to perform the adaptation at smallest sampling rate. We present an analysis of the convergence of the proposed adaptive algorithm, from which we can estimate the convergence rate and the theoretical MSE. Computational simulations were performed showing that a significant improvement in the convergence speed can be achieved for coloured signals with the proposed algorithm in comparison to a conventional full-band LMS algorithm.

Chapter 6
Blind Separation of Audio Signals

6.1 Introduction

This chapter presents a discussion on the background literature of blind separation of audio signals. Blind source separation algorithms with multiple bands have been proposed with the aim of reducing the computational cost and increasing the convergence speed of adaptation, but neglecting the effects of the spectral overlap and maintaining an efficient samples to evaluate the signals in these multiple bands [**124**], [**125**] and [**126**].

Our proposal is to deal with the problem of convolutive and determined time-domain mixtures by decomposing the signals into multiple bands using maximally decimated filter banks with real coefficients. Through the decomposition and decimation of the observed signals, we aim to increase the convergence and reduce the computational complexity of the structure; in addition to these two advantages, the use of filters with real coefficients will add another advantage for using our structure in DSP (Digital Signal Processing) implementations. One challenge will be to remove the permutation between various bands to recompose the estimated signal from the sources correctly and without degradation of the performance.

For real-world situations, we can consider audio signals as the original sources and a room or auditorium as a mixing system. The signals captured by the microphones are affected by the reverberation environment. Fig. 6.1 shows the multiple inputs and outputs system for blind source separation [**7**]. Considering that the mixtures are designed using FIR filters; we can write

$$x_m(k) = \sum_{n=1}^{N} \sum_{i=0}^{L-1} s_{mn}(i) u_n(k - i) \qquad (6.1)$$

Where N is number of sources, the FIR filters' length is L, M is the number of microphones, m=1,…,M, and the filter s_{mn} models the acoustic transfer function between the m^{th} sensor and the n^{th} source.

An adaptive algorithm is employed to estimate the separation FIR filters' coefficients ω_{nm} of U length, which makes $y_n(k)$ maximally independent. In this way, we take out the signals that are observed by the microphones (mixtures). Therefore, the output signals $y_n(k)$ are estimated by the n^{th} output described as follows:

$$y_n(k) = \sum_{m=1}^{M} \sum_{i=0}^{U-1} \omega_{nm}(i) x_m(k-i) \qquad 6.2)$$

where $n = 1,\ldots,N$.

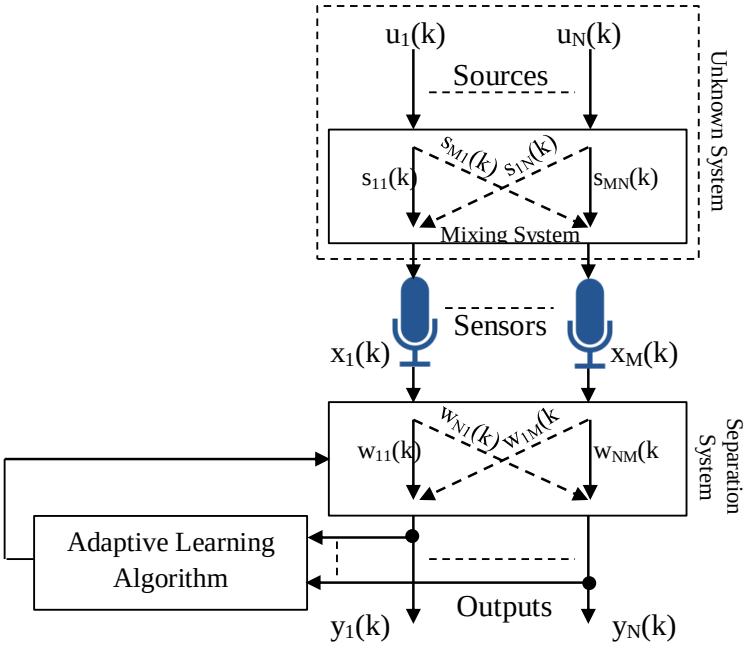


Fig. 6.1: A linear system for separating multiple inputs and outputs

6.1.1 *BSS of Determined Instantaneous Mixtures*

A special case of the separation ω_{nm} and mixing s_{mn} filters' coefficients is when both, U and L are equal to one, which means neither reverberation nor delay exist, only attenuation. Thus, our problem is reduced to the instantaneous mixture case [**17**]. Assuming the restriction $M = N$, we will have the case of determined instantaneous

mixture for which very good results can be obtained for separating mixed signals. In this case the mixing system can be described by the equations:

$$\begin{aligned} x_1(k) &= s_{11}u_1(k) + s_{12}u_2(k) + \mathrm{K} + s_{1M}u_M(k) \\ x_2(k) &= s_{21}u_1(k) + s_{22}u_2(k) + \mathrm{K} + s_{2M}u_M(k) \\ &\mathrm{M} \\ x_M(k) &= s_{M1}s_1(k) + s_{M2}u_2(k) + \mathrm{K} + s_{MM}u_M(k) \end{aligned} \tag{6.3}$$

and the system that separates the mixture (separation system) is expressed as

$$\begin{aligned} y_1(k) &= \omega_{11}x_1(k) + \omega_{12}x_2(k) + \mathrm{K} + \omega_{1M}x_M(k) \\ y_2(k) &= \omega_{21}x_1(k) + \omega_{22}x_2(k) + \mathrm{K} + \omega_{2M}x_M(k) \\ &\mathrm{M} \\ y_M(k) &= \omega_{M1}x_1(k) + \omega_{M2}x_2(k) + \mathrm{K} + \omega_{MM}x_M(k) \end{aligned} \tag{6.4}$$

Rewriting equations (6.3) in a matrix form, we have

$$x(k) = Su(k) \tag{6.5}$$

where $\boldsymbol{x}(k) = [x_1(k)\ x_2(k). \ . \ . \ x_M(k)]^T$ is the vector containing the k^{th} sample of the observed signals, $\boldsymbol{u}(k) = [u_1(k)\ u_2(k). \ . \ . \ u_M(k)]^T$ is the vector containing the signals of independent sources, the matrix $\boldsymbol{S}$ formed by the mixing coefficients s_{mn}. From equations (6.4) we have:

$$y(k) = Wx(k) \tag{6.6}$$

where $\boldsymbol{y}(k) = [y_1(k)\ y_2(k). \ . \ . \ y_M(k)]^T$ is the vector that consists of the matrix $\boldsymbol{W}$ that consists of the coefficients ω_{nm} to separate the mixture and the sources estimates $u_m(k)$. Substituting equation (6.5) in equation (6.6) will yield

$$y(k) = WSu(k), \tag{6.7}$$

which means that a matrix $\boldsymbol{W} = \boldsymbol{S}^{-1}$ exists, if the mixing matrix is non-singular, and thus the mixtures can be separated. The challenge of BSS is to accurately estimate this matrix $\boldsymbol{W}$ from the sensors. However, there are some inherent problems in BSS, known as permutation and scaling. These phenomena appear when $\mathbf{C} = \boldsymbol{WS} \neq \boldsymbol{I}$. Thus, the learning algorithm is blind, i.e. devoid of any knowledge of the sources and mixing system, it only ensures that

$$y_n(k) = c_n u_m(k), \tag{6.8}$$

where c_n is a constant value[10], other than unity, which indicates scaling of the source, and $n \neq m$ indicates existence of a permutation. The scaling and permutation effects are easily avoided with instantaneous mixtures but may limit the performance of the algorithm for convolutive mixtures as will be seen in Section 6.2.

6.1.2 *ICA*

The independent component analysis principle is taken from the neural networks [**127**], [**128**], [**129**], [**130**], where the separation matrix $\boldsymbol{W}$ was seen as a linear network. These techniques have been widely used to promote blind source separation of audio signals that are captured by multiple sensors (microphones), mixed signals and interference between signals received by a mobile station [**131**]. Some authors use the terms ICA and blind source separation interchangeably. Because of the diversity of its applications, ICA has become a good tool in various applications, such as signal processing, wireless communications, information theory, neural networks, etc.

ICA estimates the $\boldsymbol{W}$ system which extracts the mixture and sources $u_\ell(k)$ from the observed signals $x_i(k)$. The concept of independence, mentioned above, is stronger than the concept of decorrelation because while the latter involves second-order statistics, the former involves statistics of orders higher than two, which explores information that is not contained in the correlation matrix. Assuming a problem of nonlinear decorrelation of a 2×2 multiple inputs and multiple outputs-system with no loss of generality if the separation matrix is of the type $\mathbf{C} = \boldsymbol{WS} = \boldsymbol{I}$, the nonlinear function $\Psi(.)$ is an odd function so that $\Psi(y_1(k))$ also has mean zero and $y_1(k) = a$ and $y_2(k) = b$ are independent; they have zero mean [**132**], thus

$$E[\Psi(a)b] = E[\Psi(a)]E[b] = 0 \tag{6.9}$$

The choice of nonlinear function is based on theories developed for ICA, such as minimisation of mutual information, maximising non-gaussianity and maximising the

[10] For all sources to be recovered it is necessary that $c_n \neq 0$ for all n.

likelihood. The mutual information is a measure of statistical independence based on the information theory, which is always non-negative. The minimisation of mutual information (or entropy maximisation) can be interpreted as maximising the independence between components. The second approach is based on maximising the non-Gaussian (or maximisation of Kurtosis). In the theory of ICA, the random variables have non-Gaussian distributions, and this hypothesis is valid for audio signals that have typically super-Gaussian distributions, i.e. its probability density function (pdf) is large. A traditional way to evaluate the distribution of a random variable is through its kurtosis, defined as:

$$Kurtosis(x) = E\{x^4\} - 3\left(E\{x^2\}\right)^2. \tag{6.10}$$

For super-Gaussian distributions $Kurtosis(x) > 0$ for Gaussian $Kurtosis(x) = 0$ and sub-Gaussian $Kurtosis(x) < 0$.

The third method uses maximising likelihood (ML) technique. This technique is the most popular technique used for ICA. It is interesting to notice that, in fact, the three approaches described above are equivalent [**129**], [**131**] and [**133**]. Omitting the index (k) to simplify the notation, the mutual information $I(y_1, y_2)$ between the outputs $y_1(k)$ and $y_2(k)$ is expressed as

$$I(y_1, y_2) = P(y_1) + P(y_2) - P(y_1, y_2) \tag{6.11}$$

where $P(y_i)$ is the marginal entropy and $P(y_1, y_2)$ is the joint entropy of the output.

We notice from equation (6.11) that the mutual information $I(y_1, y_2)$ is minimised when the first two terms are minimised or when the third term is maximised. Gaussian signals maximise the first two terms; thus maximising the non-Gaussian leads to minimisation of mutual information. On the other hand, maximising joint output entropy will maximise the last term of equation (6.11).

To find the $\boldsymbol{W}$ in equation (6.6) that minimises the mutual information, maximises the non-gaussianity or the output likelihood, we can use the gradient method [**1**]. Considering that ($m(y)$) is the *pdf* of y, a very simple algorithm was derived in [**134**], the entropy given by

$$P(y) = -\int_{\infty}^{\infty} m(y) ln(m(y)) dy. \tag{6.12}$$

Initially, assuming a system of type $y = \omega x$ (with one input and one output), in which a nonlinear transformation applied to the output $y(\psi(y))$ is monotonically increasing or decreasing [**14**] (i.e., has a unique inverse) of the *pdf* output can be written as:

$$m(y) = \frac{m(x)}{\left|\partial y/\partial x\right|} \tag{6.13}$$

where $m(x)$ is the pdf input

Substituting equation (6.13) in (6.12):

$$P(y) = E\left[ln\left|\frac{\partial y}{\partial x}\right|\right] - E\left[ln(m(x))\right] \tag{6.14}$$

As the second term is not affected by the parameter ω, to maximise the entropy of y, it is necessary to maximise only the first term of equation (6.14). This can be done by calculating the gradient of entropy with respect to the coefficient ω, i.e.,

$$\nabla_{\omega} P \approx \frac{\partial P}{\partial \omega} = \frac{\partial}{\partial \omega}\left(ln\left|\frac{\partial y}{\partial x}\right|\right) = \left(\frac{\partial y}{\partial x}\right)^{-1} \frac{\partial}{\partial \omega}\left(\frac{\partial y}{\partial x}\right) \tag{6.15}$$

For audio signals, a nonlinear function widely used, $y = \psi(y) = \tanh(y)$, for which we have

$$\begin{aligned} \frac{\partial \psi(y)}{\partial x} &= \frac{\partial\left(\dfrac{e^{y} - e^{-y}}{e^{y} + e^{-y}}\right)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{e^{\omega x}}{e^{\omega x} + e^{-\omega x}}\right) - \frac{\partial y}{\partial x}\left(\frac{e^{-\omega x}}{e^{\omega} x + e^{-\omega x}}\right) \\ &= \frac{\omega\left(e^{\omega x} + e^{-\omega x}\right)^2}{\left(e^{\omega x} + e^{-\omega x}\right)^2} - \frac{\omega\left(e^{\omega x} - e^{-\omega x}\right)^2}{\left(e^{\omega x} + e^{-\omega x}\right)^2} = \omega\left(1 - \psi^2(y)\right) \end{aligned} \tag{6.16}$$

and

$$\frac{\partial}{\partial \omega}\left(\frac{\partial \psi y}{\partial x}\right) = 1 - \frac{\partial}{\partial \omega}\left(\omega \psi^2(y)\right) = 1 - \psi^2(y) - 2\omega\omega\psi(y\frac{\partial \psi(y)}{\partial \omega} \tag{6.17}$$

$$= 1 - \psi^2(y) - 2x\psi x\psi(y\left(1 - \psi^2(y)\right)$$

$$= (1 - \psi^2(y)) - 2x\psi x\psi(y\left(1 - \psi^2(y)\right)$$

Substituting these results into equation (6.15), the gradient of entropy is given by

$$\nabla_{\omega} P \approx \frac{1}{\omega} - 2x\psi(y) \tag{6.18}$$

Generalising to a system of M inputs and N outputs, we have the stochastic gradient of the objective function (entropy) given by

$$\nabla_{W} P \approx \left[W^{H}\right]^{-1} - 2E\left[\psi(y)x^{H}\right] \tag{6.19}$$

To increase the stability and the convergence speed of this algorithm and avoid problems of ill-conditioning of the mixing matrix, the natural gradient method was proposed in [**96**], which can be viewed as a special case of nonlinear decorrelation. The natural gradient (NG) is obtained from the traditional gradient, through the relation

$$\nabla_{W}^{NG} P = \nabla_{W} P W^{H} W \tag{6.20}$$

Substituting equation (6.19) in equation (6.20), we obtain

$$\nabla_{W}^{NG} P = \left\{I - 2E\left[\psi(y)y^{H}\right]\right\}W \,. \tag{6.21}$$

Then, using the natural gradient method, the update equation is expressed as follows:

$$W(k+1) = W(k) + \gamma\left\{I - 2E\left[\psi\left(y(k)y^{H}(k)\right)\right]\right\}W(k) \tag{6.22}$$

6.2 BSS for Convolutive Mixtures

For sources of convolutive mixtures, the algorithms described in the previous section do not perform well due to the reverberant environment. A simple way to improve source separation is to diagonalise the correlation matrix of outputs $\boldsymbol{R}_{YY}$, which for a multiple inputs and outputs linear system $M \times N$ with $M = N$ is given by

$$R_{YY} = \begin{bmatrix} \langle \Psi(Y_1) Y_1^H \rangle & \langle \Psi(Y_1) Y_2^H \rangle & \text{K} & \langle \Psi(Y_1) Y_M^H \rangle \\ \langle \Psi(Y_2) Y_1^H \rangle & \langle \Psi(Y_2) Y_2^H \rangle & \text{K} & \langle \Psi(Y_2) Y_M^H \rangle \\ \text{M} & \text{O} & \text{O} & \text{M} \\ \text{M} & \text{O} & \text{O} & \text{M} \\ \langle \Psi(Y_M) Y_1^H \rangle & \langle \Psi(Y_M) Y_2^H \rangle & \text{K} & \langle \Psi(Y_M) Y_M^H \rangle \end{bmatrix} \tag{6.23}$$

where $\langle . \rangle$ is the statistical average operator. The filters' coefficients, $\omega_{nm}(k)$, that dissolve the mixture must converge to values that minimise the mutual information between outputs, which correspond to elements that are outside the main diagonal of the correlation matrix, i.e.

$$\langle \Psi(Y_r) Y_c^H \rangle = 0 \tag{6.24}$$

where $r \neq c$.

The main diagonal elements, which control the scaling of the outputs, must be restricted to appropriate constants c_r, i.e.:

$$\langle \Psi(Y_r) Y_r^H \rangle = c_r \tag{6.25}$$

The iterative equation for updating the filter coefficients based on the method of separation of the gradient is given by

$$W_{r+1} = W_r + \gamma dW_r \tag{6.26}$$

where

$$dW = \begin{bmatrix} c_1 - \langle \Psi(Y_1) Y_1^H \rangle & \langle \Psi(Y_1) Y_2^H \rangle & \text{K} & \langle \Psi(Y_1) Y_M^H \rangle \\ \langle \Psi(Y_2) Y_1^H \rangle & c_2 - \langle \Psi(Y_2) Y_2^H \rangle & \text{K} & \langle \Psi(Y_2) Y_M^H \rangle \\ \text{M} & \text{O} & \text{O} & \text{M} \\ \text{M} & \text{O} & \text{O} & \text{M} \\ \langle \Psi(Y_M) Y_1^H \rangle & \langle \Psi(Y_M) Y_2^H \rangle & \text{K} & c_M - \langle \Psi(Y_M) Y_M^H \rangle \end{bmatrix} \tag{6.27}$$

In the procedure described above, if $\Psi(Y_r) = Y_r$ then we have a simple case of decorrelation that is insufficient to ensure the independence of the output signals of the blind source separation. However, if sources are not stationary, we can use second order statistics (SOS) considering several blocks of samples of the output signals. This method

is known as a non-stationary decorrelation (ND) [**135**]. Another method for coloured sources, which also uses SOS, considers *Time-Delayed Decorrelation* (TDD) block signals, i.e.

$$\left\langle \Psi(Y_r)Y_c^H \right\rangle = \left\langle Y_r(l)Y_c(l+\tau_r)^H \right\rangle = 0 \quad . \tag{6.28}$$

These types of decorrelation (ND and TDD) have sufficient information for estimating the separating filter; it is not necessary statistical information of higher orders to ensure the independence between the estimated samples of the sources [**131**].

On the other hand, when we consider $\Psi(Y_r) = \tanh(Y_r)$ we have

$$\left\langle \Psi(Y_r)Y_c^H \right\rangle = \left\langle \tanh(Y_r)Y_c^H \right\rangle = 0, \tag{6.29}$$

which can be seen as a case of nonlinear decorrelation, as shown in the Taylor expansion of tanh(Y_r), given by

$$\left\langle \left(Y_r - \frac{Y_r^3}{3} + \frac{2Y_r^5}{15} - \frac{17Y_r^7}{315} + \mathrm{K} \right) Y_2^H \right\rangle = 0 \tag{6.30}$$

Therefore, this method uses higher-order statistics (HOS) or nonlinear decorrelation and can solve the problem of blind separation of convolutive mixtures.

6.2.1 ***BSS in TD***

This section discusses the full-band time domain approach and our novel normalised approach. The BSS is based on the hypothesis that states that different signal sources are statistically mutually independent. In the real world, due to reverberant environment, the signals of the original sources are filtered by a linear system with multiple inputs and outputs before being captured by the microphones. From this moment, we assume that the number of sources is equal to the number of microphones as shown in

Fig. 6.2 [**136**].

In a blind source separation problem, we are interested in the system that separates the mixture; in this case, it is described by

$$y_n(k) = \sum_{m=1}^{M} \sum_{i=0}^{U-1} \tilde{\omega}_{mn}(i) x_m(k-i). \tag{6.31}$$

The following extension will use the coefficients $\tilde{\omega}_{mn}(i)$ as $\tilde{\omega}_{mn}(i) = \omega_{nm}(i)$. Extending the formulation of the output signals to a matrix form, we can describe the n^{th} output signal at time k as

$$y_n(k) = \sum_{m=1}^{M} x_m^T(k) \tilde{\omega}_{mn} \tag{6.32}$$

where $\boldsymbol{x}_m(k) = [x_m(k), x_m(k\text{-}1), \ldots, x_m(k - U + 1)]^T$ has the newest U samples taken by the m^{th} microphone and $\tilde{\omega}_{mn}(n) = [\tilde{\omega}_{mn}(0), \tilde{\omega}_{mn}(1), \ldots, \tilde{\omega}_{mn}(U-1)]^T$ is the vector that consists of U coefficients of FIR which reflects the way from sensor m to the output n.

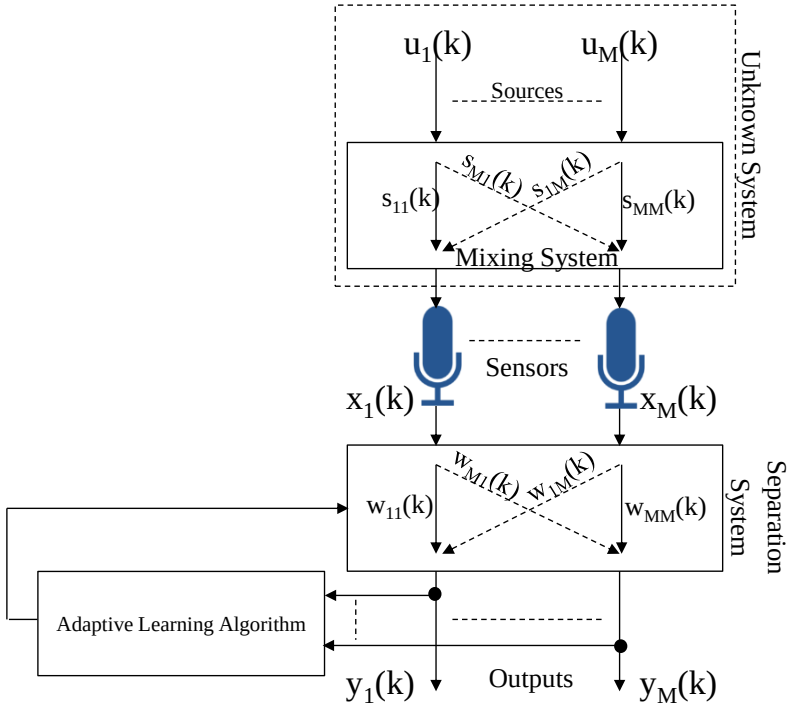


Fig. 6.2: Linear system with *M* inputs and *M* outputs a Convolutive mixture.

Two new parameters needed for generalisation of the formulation are the number of time-delays D taken into account in calculating the correlation ($1 \leq D \leq U$) and size K of the output signal block.

From equation (6.32), we can describe *K-samples* at the n^{th} output and at time l as follows:

$$y_n(l) = \sum_{m=1}^{M} \tilde{X}_{K\,m}^{\;T}(l)\tilde{\omega}_{mn} = \left[y_n(lU) \quad \Lambda \quad y_n(lU + K - 1)\right]^T \tag{6.33}$$

Now, we will build $\tilde{X}_m(l)$ in a Toeplitz matrix form, based on $\tilde{X}_{K\,m}(l)$, of dimension $U \times K$, containing U blocks with delayed versions of the samples of the signal captured by the m^{th} sensor, we obtain:

$$\widetilde{X}_m(l) = \begin{bmatrix} x_m(lU) & \Lambda & x_m(lU+U-1) \\ x_m(lU+1) & \mathrm{O} & x_m(lU-U-2) \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ x_m(lU+K-1) & \Lambda & x_m(lU-U-K) \end{bmatrix} \tag{6.34}$$

Then equation (6.33) can be extended to include samples of D blocks delayed in time. Therefore, the matrix with the n^{th} output, of dimension $K \times D$, can be expressed as:

$$Y_n(l) = \sum_{m=1}^{M} X_m^T(l)\widetilde{W}_{mn} \tag{6.35}$$

$$X_m(l) = \left[\widetilde{X}_m^T(l), \widetilde{X}_m^T(l-1)\right] \tag{6.36}$$

To ensure the linear convolution of $\boldsymbol{Y}_n(l)$ to the maximum number of time delay, $D = U$ [**136**], it takes two blocks of input signals $\widetilde{\boldsymbol{X}}_m^T$. Therefore, the dimensions of $\boldsymbol{X}_m(l)$ and $\widetilde{\boldsymbol{W}}_{mn}$ are $K \times 2U$ and $2U$ x D, respectively. The matrices $\boldsymbol{X}_m(l)$ are obtained by doubling the array size $\widetilde{\boldsymbol{X}}_m$:

$$X_m(l) = \left[\widetilde{X}_m^T(l), \widetilde{X}_m^T(l-1)\right] \tag{6.37}$$

The matrix $\widetilde{\boldsymbol{X}}_m^T(l-1)$ is also a Toeplitz matrix, so that the first row of the matrix $\boldsymbol{X}_m(l)$ contains $2U$ samples of m^{th} input signal and each subsequent row is obtained by shifting the previous row to the right by one sample. The matrix $\widetilde{\boldsymbol{W}}_{mn}$ is Sylvester matrix type of dimension $2U$ x D, defined as

$$\tilde{W}_{mn} = \begin{bmatrix} \tilde{\omega}_{mn}(0) & 0 & \text{K} & 0 \\ \tilde{\omega}_{mn}(1) & \tilde{\omega}_{mn}(0) & \text{O} & \text{M} \\ \text{M} & \tilde{\omega}_{mn}(1) & \text{O} & 0 \\ \tilde{\omega}_{mn}(U-1) & \text{M} & \text{O} & \tilde{\omega}_{mn}(0) \\ 0 & \tilde{\omega}_{mn}(U-1) & \text{O} & \tilde{\omega}_{mn}(1) \\ \text{M} & 0 & \text{O} & \text{M} \\ 0 & \text{K} & 0 & \tilde{\omega}_{mn}(U-1) \\ 0 & \text{K} & 0 & 0 \\ \text{M} & \text{K} & \text{O} & \text{M} \\ 0 & \text{K} & 0 & 0 \end{bmatrix} \tag{6.38}$$

which is the latest U - D + 1 rows formed by zeros to ensure compatibility with $\boldsymbol{X}_m(l)$ for general case $1 \leq D \leq U$. To allow a convenient notation update algorithm, which will be seen below, we can rewrite equation (6.35) for a more compact form, i.e.

$$Y(l) = X(l)\tilde{W} \tag{6.39}$$

where

$$Y(l) = \begin{bmatrix} Y_1(l) & \Lambda & Y_M(l) \end{bmatrix} \tag{6.40}$$

is a matrix of dimension $K \times M\,D$ containing the building blocks of the output signals of all bands;

$$X(l) = \begin{bmatrix} X_1(l) & \Lambda & X_M(l) \end{bmatrix} \tag{6.41}$$

is a matrix of order $K \times 2UM$ containing all time-delayed blocks of all sensors, and

$$\tilde{W} = \begin{bmatrix} \tilde{W}_{11} & \Lambda & \tilde{W}_{1M} \\ \text{M} & \text{O} & \text{M} \\ \tilde{W}_{M1} & \Lambda & \tilde{W}_{MM} \end{bmatrix} \tag{6.42}$$

is a matrix of dimension $2UM$ x DM containing all coefficients of all separation filters.

A scheme that employs filter banks of UEPS with complex coefficients applied to the blind source separation that was used in [**137**] and [**138**].

6.2.1.1 ***Objective Function and Algorithm Update***

The expansion below uses $\tilde{s}_{nm}(i)$ coefficients, where $\tilde{s}_{nm}(i) = s_{mn}(i)$. Similar to the separation system described by equation (6.39), the mixing system can be modelled by $\boldsymbol{X}(l) = \boldsymbol{U}(l)\tilde{\mathbf{S}}$, where $\boldsymbol{U}(l)$ is a matrix of $K \times M(L+U-1)$ dimension containing the delayed versions of the source signals, and $\tilde{\mathbf{S}}$ is the mixing matrix of Sylvester type of order $M(L + U - 1) \times 2MU$ containing the coefficients of the impulse response of all $\tilde{s}_{nm}(k)$ filters. These dimensions result, again, from the performed condition of convolution linearity. It is therefore possible to obtain a *block-diagonal* (*bdiag*) matrix $\mathbf{C} = \tilde{\mathbf{S}}\widetilde{\mathbf{W}}$, such that $\mathbf{C} - bdiag\ \mathbf{C} = boff\ \mathbf{C} = 0$.

The *bdiag* operator operates on a matrix formed by submatrices, zeroing all submatrices that do not belong to the main diagonal. Similarly, the *block-off-diagonal* (*boff*) operator resets all submatrices located on the off-diagonals. To achieve this goal, using second-order statistics, a series of correlation matrices will describe the signals, with time delays based on the matrix formulation presented earlier. These matrices are defined by

$$R_{xx}(l) = X^{H}(l)X(l) \tag{6.43}$$

and

$$R_{yy}(l) = Y^{H}(l)Y(l) \tag{6.44}$$

with dimensions $2MU$ x $2MU$ and MD x MD, respectively. For equation (6.44) to have full rank, it is necessary that the length of the block output vector to be $K \geq D$.

The objective function, based on the sequence of delayed correlation matrices, is given by [**53**]

$$\varsigma(l) = \sum_{r=0}^{\infty} \alpha(r,l)\left\{log\left[det\left(bdiag\left(Y^{H}(r)Y(r)\right)\right)\right] - log\left(det\left(Y^{H}(r)Y(r)\right)\right)\right\} \tag{6.45}$$

whose equilibrium point solution of [**139**] corresponds exactly to the desired blind source separation, or $boff\ \mathbf{C} = \mathbf{0}$. The parameter α is a constant normalised according to $\sum_{r=0}^{\infty} \alpha(r,l) = 1$. Using the matrix formulation of equation (6.39) to calculate the

matrices of the minimised temporal correlation of equation (6.44), the objective function contains D time delays of autocorrelations and cross-correlations of output signals from the blind source separation.

Considering an algorithm based on the gradient method, the recursive equation for updating the coefficients of the filters that extract the mixture is given by

$$\widetilde{W}(l+1) = \widetilde{W}(l) - \gamma \nabla_{\widetilde{W}} \varsigma(l) \tag{6.46}$$

Using the formulation of the natural gradient [**140**] which is more robust and the computationally complexity is reduced; we obtain the following recursive equation for updating the coefficients:

$$\widetilde{W}(l+1) = \widetilde{W}(l) - \gamma \nabla^{NG}_{\widetilde{W}} \varsigma(l) \tag{6.47}$$

where the natural gradient of the objection function (equation (6.45)) can be expressed as:

$$\nabla^{NG}_{\widetilde{W}} \varsigma(l) = \widetilde{W}\widetilde{W}^{H} \nabla_{\widetilde{W}} \varsigma(l) \tag{6.48}$$

$$= 2\sum_{r=0}^{\infty} \alpha(r,l)\widetilde{W}\left\{R_{yy}(r) - bdiagR_{yy}(r)\right\}bdiag^{-1}R_{yy}(r) \tag{6.49}$$

and γ is the learning rate of the algorithm, where $bdiag(\cdot)$ operator interprets the matrix to which it is applied as a composition of submatrices, zeroing all submatrices that do not belong to its main diagonal. To illustrate this operator, consider a system with 3 sources. In this case, the matrix $\mathbf{R_{yy}}(l)$ can be described as:

$$R_{yy}(l) = \begin{bmatrix} Y_1^H(l)Y_1(l) & Y_1^H(l)Y_2(l) & Y_1^H(l)Y_3(l) \\ Y_2^H(l)Y_1(l) & Y_2^H(l)Y_2(l) & Y_2^H(l)Y_3(l) \\ Y_3^H(l)Y_1(l) & Y_3^H(l)Y_2(l) & Y_3^H(l)Y_3(l) \end{bmatrix} \tag{6.50}$$

The matrices $\mathbf{Y}_r^H(l)\mathbf{Y}_r(l)$ (with r = 1, 2 and 3) are the autocorrelation matrices of the r^{th} output, while the matrices $\mathbf{Y}_r^H(l)\mathbf{Y}_c(l)$, with $r \neq c$, are the matrices of cross-correlation between the r^{th} and c^{th} output. It is normal to subdivide $\mathbf{R_{yy}}(l)$ matrix into submatrices, and the autocorrelation submatrices belonging to the main diagonal of the matrix of submatrices. So bdiag $\mathbf{R_{yy}}(l)$ yields the following result:

$$bdiagR_{yy}(l) = \begin{bmatrix} Y_1^H(l)Y_1(l) & 0 & 0 \\ 0 & Y_2^H(l)Y_2(l) & 0 \\ 0 & 0 & Y_3^H(l)Y_3(l) \end{bmatrix} \quad (6.51)$$

where 0 is a zero matrix with the same dimensions as $\boldsymbol{Y}_r^H(l)\boldsymbol{Y}_c(l) = \boldsymbol{R}_{\boldsymbol{y}_r\boldsymbol{y}_c}(l)$.

During the process of updating the coefficients, it is necessary to ensure the structure of the Sylvester matrix $\widetilde{\boldsymbol{W}}(l+1)$. The indiscriminate use of a gradient that acts on the entire matrix can destroy this feature by removing the redundancy that allows a two-way relationship between the matrices $\widetilde{\boldsymbol{W}}_{mn}$ and the corresponding filters($\widetilde{\mathbf{w}}_{mn}$). This is easily imposed by selecting one of the columns of matrices $\widetilde{\boldsymbol{W}}_{mn}$ which contains all coefficients of the filters $\widetilde{\omega}_{mn}(i)$ (for i = 0, ..., U - 1) and generates $\nabla_{\widetilde{W}}^{NG}\ \zeta(l)$ according to equation (6.38). In [**53**] it is shown that the choice of the first U elements of column $\widetilde{\boldsymbol{W}}_{mn}$ is the best choice for optimisation purposes. The algorithm that processes the signal to generate an iteration to adjust the coefficients, considering a system with two sources and two sensors (two inputs and two outputs), is given by:

$$\widetilde{W}(r) = \widetilde{W}(r-1) = \begin{bmatrix} \widetilde{W}_{12}R_{y_2y_1}R_{y_1y_1}^{-1} & \widetilde{W}_{11}R_{y_1y_2}R_{y_2y_2}^{-1} \\ \widetilde{W}_{22}R_{y_2y_1}R_{y_1y_1}^{-1} & \widetilde{W}_{21}R_{y_1y_2}R_{y_2y_2}^{-1} \end{bmatrix} \quad (6.52)$$

where

$$R_{nn}(l) \approx diag\{R_{nn}(l)\} \quad (6.53)$$

Thus, $\mathbf{R}_{nn}^{-1}$ (equation (6.52)) can be obtained by inverting each element of its diagonal, which corresponds to the inverse of the power of the blocks of the delayed-time of the n^{th} output signals (see equation (6.33)).

6.2.1.2 *Computational Cost*

We consider computational cost as the actual number of multiplications per block of the input signal (N_{block}) necessary to update the coefficients of FIR separation filters, considering the direct form implementations.

The total $N_{block,FB}$ required for full-band (FB) scheme [**53**], considering a linear system $M \times$ M (M sensors and M sources), is given by

$$N_{blockFB} = 2M^2KUD + M^2(K + DK) + M^2\left(\left(\sum_{c=1}^{U} c\right) + U\right) \tag{6.54}$$

The first term relates to the cost of generating all the delay-time blocks for all outputs (equation (6.39)). The second term is the cost of obtaining the submatrices $\mathbf{R}_{y_n y_m}(m)$ of the matrix $\mathbf{R}_{yy}$ (see equation (6.52)) for calculating the gradient. The third term indicates the cost of adaptation considering only the update the first U rows of the first column of the matrix $\widetilde{\mathbf{W}}_{mn}(l)$ (equation (6.38)).

6.2.2 ***BSS in FD***

The blind source separation method in the frequency domain aims to transform the convolutive mixture in the time domain into an instantaneous mixture problem in the frequency domain, by applying the solution shown in equation (6.22) (adapted to complex values) in each frequency band and resolving the blind source separation [**95**], [**140**], [**141**]. Therefore, the signals captured by microphones in the transformed domain can be considered linear mixtures, i.e.:

$$X(\omega) = S(\omega)U(\omega) \tag{6.55}$$

where

$$X(\omega) = \begin{bmatrix} X_1(\omega) & \Lambda & X_M(\omega) \end{bmatrix}^T \tag{6.56}$$

is the vector that contains the *Fourier* transform of M observed signals,

$$U(\omega) = \begin{bmatrix} U_1(\omega) & \Lambda & U_N(\omega) \end{bmatrix}^T \tag{6.57}$$

is the vector containing the *Fourier* transform of N source signals and

$$S(\omega) = \begin{bmatrix} S_{11}(\omega) & \Lambda & S_{1N}(\omega) \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ S_{M1}(\omega) & \Lambda & S_{MN}(\omega) \end{bmatrix} \tag{6.58}$$

is the matrix that contains the response of the frequency ω of FIR filters $s_{mn}(k)$ that mix the signals from all sources $u_n(k)$. Fig. 6.3 shows the overall scheme of the blind source separation, where the observed signals $x_m(k)$ in the time domain are "windowed" and converted to the $X_M(\omega, i)$ signals in the frequency domain through a STFT:

$$X_M(\omega,i) = \sum_{r=-\frac{U}{2}}^{\frac{U}{2}-1} x_m(i+r)win(r)e^{-j\Omega r}, \tag{6.59}$$

where $\Omega \in \left\{0, \frac{1}{U}2\pi, \cdots, \frac{U-1}{U}2\pi\right\}$ is the normalised frequency, win(r) function with the window edges smoothed around zero (usually Von Hann[11]) and i is the index of the sample windowed signal. The relationship between the frequency ω and a particular bin f can be expressed as follows: $\omega = \frac{2\pi f}{U}$. Following the same notation, we can express each bin of the N output time series as

$$Y(f,t) = W(f)X(f,t) \tag{6.60}$$

where

$$Y(f,t) = \left[Y_1(f,t) \quad \Lambda \quad Y_N(f,t)\right]^T \tag{6.61}$$

and

$$W(f) = \begin{bmatrix} W_{11}(f) & \mathrm{K} & W_{1M}(f) \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ W_{N1}(f) & \mathrm{K} & W_{NM}(f) \end{bmatrix} \tag{6.62}$$

According to the above equation, the convolutve mixture in equation ((6.1) is seen like [**95**] instantaneous mixture.

This procedure should be applied to all frequency bins. Finally, we apply the inverse discrete *Fourier* transform (IDFT) to obtain the coefficients of the filters $\omega_{nm}(k)$ and obtain the resulting signal in the N outputs.

[11] The performance of the BSS algorithms using Von Hann is much better than the performance of the other windows [**140**].

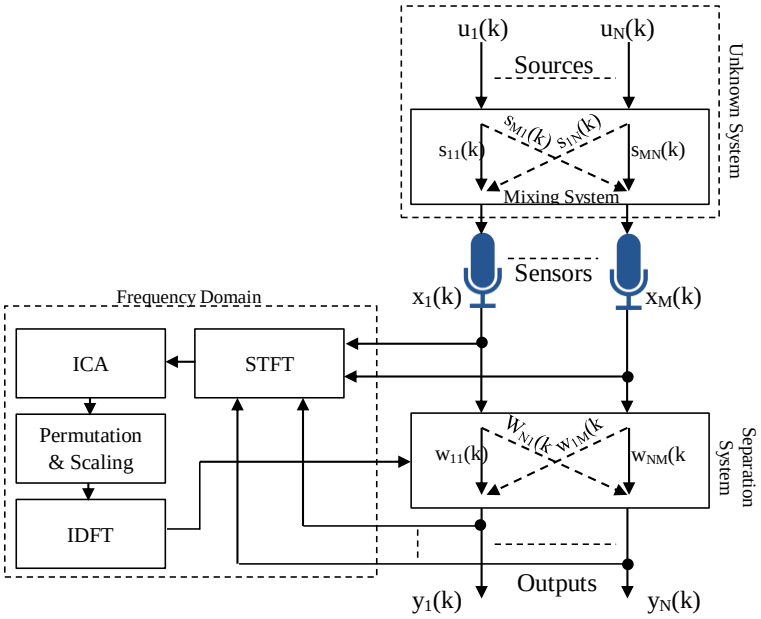


Fig. 6.3: General scheme for BSS in FD.

6.2.2.1 *Adaptive Algorithm*

The adaptation algorithm obtained for instantaneous real mixtures (6.22) can be used for the separation of convolutive mixtures in the frequency domain, by adapting it carefully for instantaneous mixtures with complex values. The update equation based on the natural gradient for each line is given by

$$W_{r+1}(f) = W_r(f) - \gamma\left[I - \left\langle 2\Psi(Y(f,t))Y^H(f,t)\right\rangle_t\right]W_r(f) \tag{6.63}$$

the operator $\langle . \rangle_t$ is the average time, the index r refers to r^{th} iteration, and γ *is* the convergence parameter,

$$\Psi(Y(f,t)) = \left[\Psi(Y_1(f,t)),\mathrm{K}\ ,\Psi(Y_N(f,t))\right]^T \tag{6.64}$$

and the nonlinear function at outputs is

$$\Psi(Y_n(f,t)) \equiv tanh(Re(Y_n(f,t))) + tanh(Im(Y_n(f,t)))\ldots \tag{6.65}$$

6.2.2.2 *Permutation and Scaling*

In blind source separation based on independent component analysis, separation filters are updated so that the output becomes mutually independent. This type of solution may have permutation between the outputs and this can be scaled by a constant factor. The solution in the frequency domain is performed directly in the bins of the DFT filters that separate the mixture. In this case, the permutation and scaling between bins of different filters leads to deterioration in the performance of the adaptive algorithm. This is the biggest challenge.

Some proposals to reduce the occurrence of permutations can be found in this area: limiting the filter's length in the time domain [**142**], [**143**], minimisation of correlation [**144**], estimating the direction of arrival for each source [**145**], and mixed solutions of the last two techniques [**146**].

To reduce the scaling ambiguity, it was proposed in [**147**] that forcing all matrices of dimension $N \times N$ to dissolve the mixture that has a determinant unit, i.e.

$$W(f)^{norm} = W(f)^{orig} \left| W(f)^{orig} \right|^{-\frac{1}{Q}} \tag{6.66}$$

where $\boldsymbol{W}(f)^{norm}$ is the normalised separation matrix to be used (with unit determinant) and $\boldsymbol{W}(f)^{orig}$ is the original matrix. This ensures the conservation of the size of the separation matrices, keeping the spectral contents unchanged. Another way to overcome the problem of scaling is by the principle of minimal distortion [**148**], which modifies the separation matrix $\boldsymbol{W}(f)$ after the method of convergence, as follows:

$$W(f) \leftarrow diag\left[W^{-1}(f)\right] W(f), \tag{6.67}$$

where the operator $diag(\cdot)$ cancels all the coefficients of a matrix that are not in its main diagonal. Assuming that the separation is reasonable and there is permutation, we can approximate the matrix:

$$W(f) \approx D(f) S^{-1}(f), \tag{6.68}$$

where $D(f)$ is a diagonal matrix that contains scaling parameters. This approach implies that the transformation in equation (6.67) generates a matrix $\mathbf{W}(f)$ with a reasonable scaling and not with a completely random scaling.

Another difficulty concerns the equivalence between the linear convolution, implemented in the time domain, and the circular convolution, implemented in the frequency domain. To circumvent this problem, a restriction is performed on the separation filters in the time domain [**149**].

6.3 Performance Measures

There are several forms of performance evaluation for the source separation methods. A widely quantitative measure that is often used is the signal to interference ratio (SIR) [**150**], defined for determined instantaneous mixtures as

$$SIR = 10log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} SIR_n\right) \tag{6.69}$$

with $SIR_n = max\left[SIR_{n1,\ldots,}SIR_{nN}\right]$ and

$$SIR_{nm} = \frac{\sum_k \left|C_{mn}S_n(k)\right|^2}{\sum_{\lambda=1,\lambda\neq n}^{N}\sum_k \left|C_{m\lambda}S_\lambda(i)\right|^2} \tag{6.70}$$

where SIR_{nm} represents the relation between the signal power present at the output $y_m(k)$ if the active $s_n(k)$ exists and the output power $y_m(k)$ if $s_n(k)$ is inactive and $C_{m\ell}$ is the element of the ℓ^{th} column and the m^{th} row of the matrix $\mathbf{C} = \mathbf{WP}$.

There are qualitative ways to evaluate the separation performance involving the visualisation of signals over time; followed by a comparison between the source signals, mixtures and obtained outputs (estimates of the sources) or, through the spectral analysis of the magnitude over the time, that is, comparing the observed signals with the original signals' spectrogram and how the estimated signals can realise the quality of separation.

A subjective form of evaluation would listen to the observed and separated signals, and observe the level of separation.

In the case of convolutive mixtures, it is necessary to make a modification to the SIR measure evaluation presented earlier. In this case, equation (6.70) should be generalised as follows:

$$SIR_{rc} = \frac{\sum_{k} \left| c_{cr}(k) * u_r(k) \right|^2}{\sum_{\lambda=1, \lambda \neq r}^{K} \sum_{k} \left| c_{c\lambda}(k) * u_\lambda(k) \right|^2} \tag{6.71}$$

where $\mathbf{c}_{c\ell}(k)$ is the sum of the convolutions of the filters. This sum is for the c^{th} row of the matrix **W** with the filters of ℓ^{th} column of the matrix **P**, i.e., **C** = **W** * **P**. The operator "*" represents the convolution operation.

For instance, for a linear system with two inputs and two outputs Fig. 6.4(*a*) shows the estimated output $y_1(k)$ when only the source $u_1(k)$ is active and Fig. 6.4(*b*) shows the estimated output of $y_2(k)$ when only the source $u_2(k)$ is active. In this case, assuming no permutation problem, we can define:

$$SIR_1 = \frac{\left(u_1(k) * \left[p_{11}(k) * \omega_{11}(k) + p_{12}(k) * \omega_{21}(k)\right]\right)^2}{\left(u_2(k) * \left[p_{21}(k) * \omega_{11}(k) + p_{22}(k) * \omega_{21}(k)\right]\right)^2} \tag{6.72}$$

$$SIR_2 = \frac{\left(u_2(k) * \left[p_{22}(k) * \omega_{22}(k) + p_{21}(k) * \omega_{12}(k)\right]\right)^2}{\left(u_1(k) * \left[p_{12}(k) * \omega_{22}(k) + p_{11}(k) * \omega_{12}(k)\right]\right)^2} \tag{6.73}$$

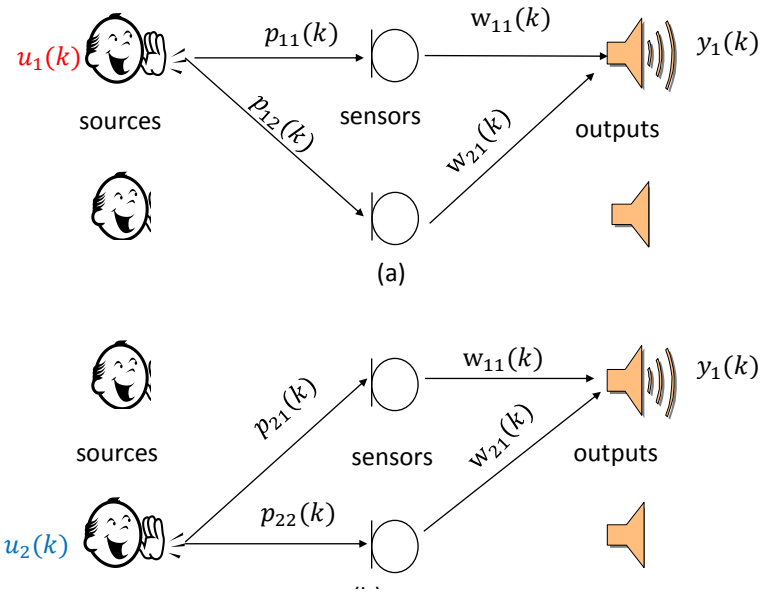


Fig. 6.4 Scheme for measuring SIR_1 in a 2-inputs/2-outputs system

6.3.1 *Global Performance Measures*

An alternative way to evaluate the overall performance of blind source separation methods was proposed in [**150**]. This method does not require knowledge of the mixing system; only the original sources and their estimates.

Assuming uncorrelated sources u_n (n = 1,..., N) and absence of measurement noise, $\tilde{u}_n$ (the estimate of the n^{th} source) can be decomposed as:

$$\tilde{u}_n = u_n^{target} + e_n^{interf} + e_n^{artef} \tag{6.74}$$

where

$$u_n^{target} = \frac{\langle \tilde{u}_n, u_n \rangle u_n}{\|u_n\|^2}, \tag{6.75}$$

is an acceptable distortion in the n^{th} source,

$$e_n^{interf} = \sum_{n' \neq n} \frac{\langle \tilde{u}_n, \tilde{u}_{n'} \rangle u_{n'}}{\|u_{n'}\|^2}, \tag{6.76}$$

is a distortion in the n^{th} source caused by the interference of other N-1 sources and

$$e_n^{artef} = \tilde{u}_n - u_n^{target} - e_n^{interf}, \tag{6.77}$$

are the artifacts introduced by separation algorithms.

After the decomposition that is shown in Fig. 6.5, we can define three global measures of performance [**151**], i.e. the Global SIR (SIR_G), the global signal-to-artifact (SAR_G) and the global signal-to-distortion ratio (SDR_G), that are given by:

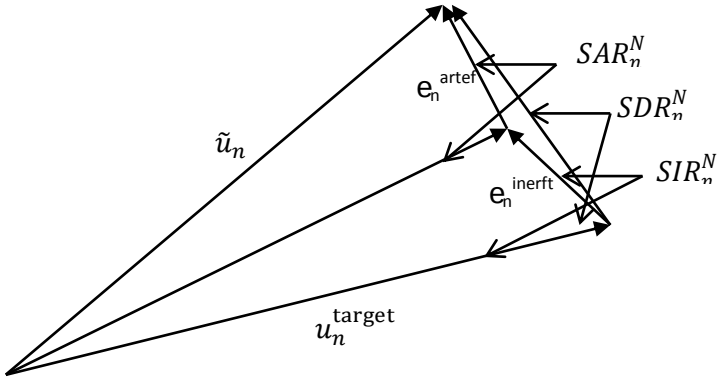


Fig. 6.5: Decomposition of the estimated n^{th} source for calculation of performance measures

$$SIR^G = 10log_{10}\frac{\left\|u_n^{target}\right\|^2}{\left\|e_n^{interf}\right\|^2}, \tag{6.78}$$

$$SAR^G = 10log_{10}\frac{\left\|u_n^{target} + e_n^{interf}\right\|^2}{\left\|e_n^{artef}\right\|}, \tag{6.79}$$

$$SDR^G = 10log_{10}\frac{\left\|u_n^{target}\right\|^2}{\left\|e_n^{interf} + e_n^{artef}\right\|^2}. \tag{6.80}$$

The global mean performance measures can be obtained as described below:

$$SIR^G = 10log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} SIR_n^G\right), \tag{6.81}$$

$$SAR^G = 10log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} SAR_n^G\right), \tag{6.82}$$

$$SDR^G = 10log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} SDR_n^G\right). \tag{6.83}$$

Chapter 7
NMD_UEPS for BSS

7.1 Introduction

In this chapter, we present our second proposal for blind separation of a convolutive mixture, using filter banks. An algorithm with a bank of subsampled filters of UEPS is used, where frequency bands are decomposed into octaves. The use of multiple bands in the separation filters results in lower sampling rates and lower orders. As a direct consequence, the proposed method has less computational cost, a greater convergence rate and a higher signal to interference ratio in the stationary state when compared to the corresponding algorithms in full-band. Filter coefficients that break down mixtures in each band are set independently by an algorithm in the time domain, which employs second-order statistics and the new normalisation scheme.

7.2 Multiband Blind Source Separation in Time Domain

We first propose a further reduction to equation (6.53) of the normalisation factor to a scalar. In this case,

$$R_{nn}(l) \approx \left(y_n(l)\right)^T y_n(l) I \tag{7.1}$$

with $y_n^i(l)$ is given by equation (6.33). This reduction will significantly decrease the computational cost.

For higher order separation filters, equation (6.54) can be simplified by considering only the dominant terms, resulting in

$$N_{blockFB} \approx 4M^2U^3 \tag{7.2}$$

The expression above was obtained by considering the normalisation factor of the simplified equation (7.1), $D = U$ and $K = 2D$.

Our proposal is to use a NMD_UEPS filter bank with octave bands. Fig. 7.1 shows a blind source separation system with two inputs and two outputs considering an UEPS filter bank with *L* bands. This is a modified version of the proposed work presented in Chapter 5 [**104**] . The signals at the separation filters inputs of each band $\widetilde{\omega}_{mn}^{i}(k)$ are decimated by half of the maximal decimation factor to minimise the alaising throughout the adaptation process of the coefficients. The output signals for the separation filter of each band are decimated by a factor of 2 to restore the maximal sampling rate of the scheme before the reconstruction stage of the output signal.

For PR in a two-channel cosine-modulated filter bank with EP, we can write (see equations (4.15) and (4.16)) [**100**]:

$$p^{\lambda}(k) = 2p_m(k)\cos\left[\frac{\pi}{2}(\lambda + 0.5)\left(k - \frac{K_M - 1}{2}\right) + \varphi_{\lambda}\right] \tag{7.3}$$

$$q^{\lambda}(k) = 2p_m(k)\cos\left[\frac{\pi}{2}(\lambda + 0.5)\left(k - \frac{K_M - 1}{2}\right) - \varphi_{\lambda}\right] \tag{7.4}$$

where $\varphi_{\ell} = (-1)^{\ell}\frac{\pi}{4}$ for $\ell = 0.1$ and $0 \leq k \leq K_M - 1$.

For a two-channel octave-bands filter bank, (equations (7.3) and (7.4)), the analysis filters ($P_i(z)$) and synthesis filters ($Q_i(z)$) are equivalent to the EP with *L*-band tree filter bank that are obtained as presented in Section 4.4. The number of coefficients of each separation filter in i^{th} band $\omega_{mn}^{i}(k)$ must be at least [**104**]

$$U_i = 2\left[\frac{U - 1 + K_{Q_i}}{\mathcal{F}_i}\right] \tag{7.5}$$

for the i^{th} synthesis filter bank, the order is K_{Q_i}.

Every separation filter can be regulated individually by its own coefficients. The full-band approach (equation (6.47)) applied to the multiple bands, and the update equation of the filter parameters of the i^{th} band is expressed as follows:

$$\widetilde{W}^{i}(r) = \widetilde{W}^{i}(r-1) - \frac{2}{b_i}\left\{\sum_{l=1}^{b_i}\begin{bmatrix} \widetilde{W}_{12}^{i}R_{21}^{i}R_{11}^{i-1} & \widetilde{W}_{11}^{i}R_{12}^{i}R_{22}^{i-1} \\ \widetilde{W}_{22}^{i}R_{21}^{i}R_{11}^{i-1} & \widetilde{W}_{21}^{i}R_{12}^{i}R_{22}^{i-1} \end{bmatrix}\right\}\begin{bmatrix} \gamma_1^i I & 0 \\ 0 & \gamma_2^i I \end{bmatrix} \quad (7.6)$$

where

$$R_{mn}^{i}(l) = \left[Y_m^i(l)\right]^H Y_n^i(l) \quad (7.7)$$

and

$$Y_n^i(l) = \begin{bmatrix} y_n^i(lU_i) & \Lambda & y_n^i(lU_i - D_i + 1) \\ y_n^i(lU_i + 1) & \mathrm{O} & y_n^i(lU_i - D_i + 2) \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ y_n^i(lU_i + K_i - 1) & \Lambda & y_n^i(lU_i - D_i + K_i) \end{bmatrix} \quad (7.8)$$

The above matrices have dimensions D_i x D_i (with $1 \leq D_i \leq U_i$) and $K_i \times D_i$ (with $K_i \geq D_i$), respectively, b_i is the number of blocks, K_i is the size of each block, and γ_n^i is the n^{th} adaptation step of the i^{th} band, and r is the number of iterations for all bands (see Fig. 7.1).

For the purpose of lower computational cost, the normalisation factor $R_{nn}^{i-1}(m)$ can be simplified to a scalar (see equation (7.1)). In this case,

$$R_{nn}^{i}(l) \approx \left[y_n^i(l)\right]^T y_n^i(l) I \quad (7.9)$$

with $\boldsymbol{y}_n^i(l)$ corresponding with the first column of the matrix of delay-time block (equation (7.8)).

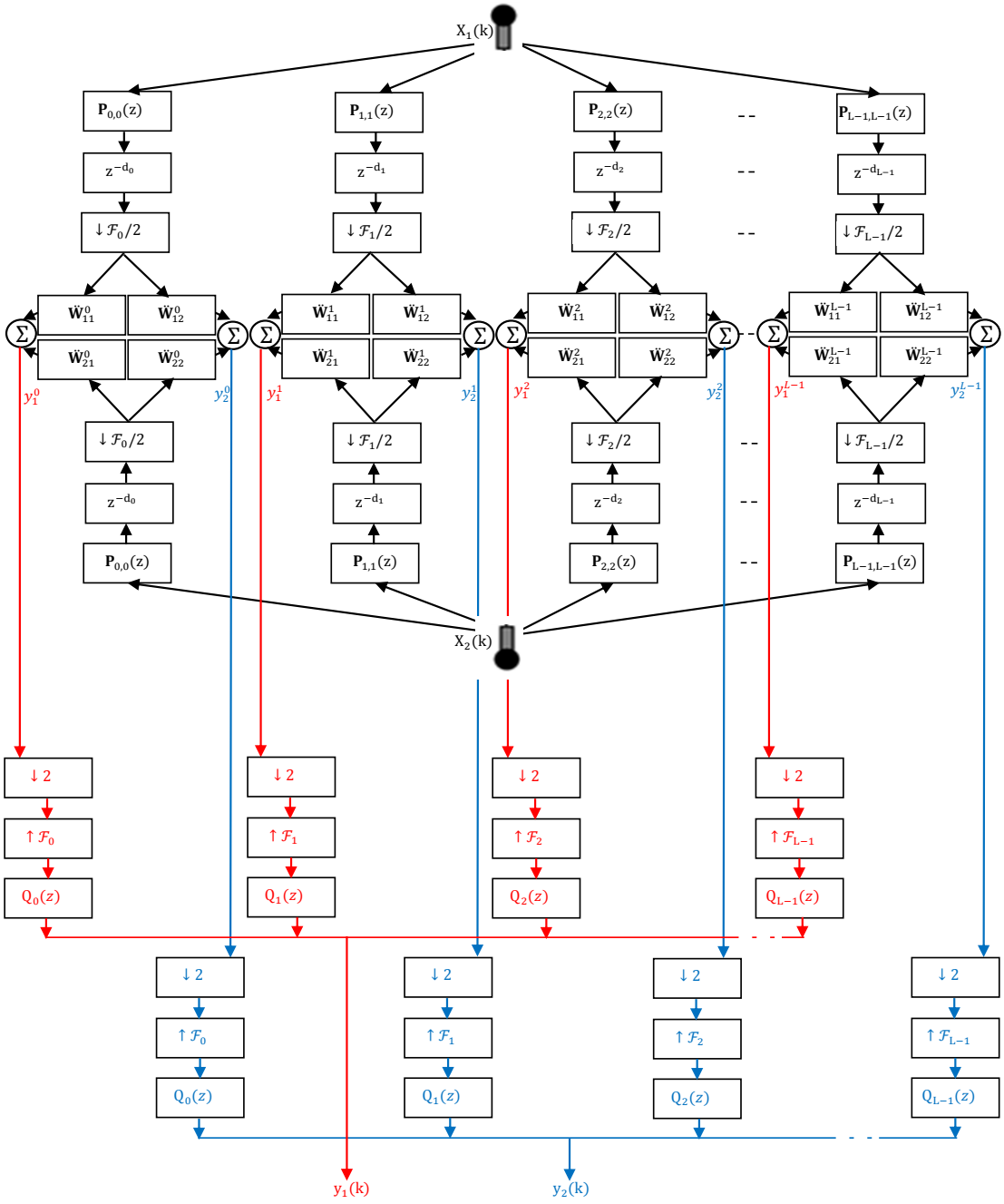


Fig. 7.1 Blind source separation scheme with two inputs and two outputs.

The implementation of the proposed multiband version is more flexible than the full-band version. For example, it can work with separation filters of different lengths and independent learning rates of various bands.

7.2.1 *Computational Cost*

The $N_{block.}$ that is required to update FIR separation filters' coefficients in L number of bands of the UEPS for an $M \times M$ linear system with multiple inputs and multiple outputs, is given by [**53**]

$$N_{blockMB} = M\sum_{i=1}^{L}\frac{K_{P_{ii}}K_i}{\frac{F_i}{2}} + M^2\sum_{i=1}^{L}\frac{2U_iD_iK_i}{\frac{F_i}{2}} + MU\sum_{i=1}^{L}\frac{K_{Q_i}}{F_i}$$
$$+ M^2\sum_{i=1}^{L}\frac{K_i + D_iK_i}{\frac{F_i}{2}} + M^2\sum_{i=1}^{L}\frac{\left(\sum_{c=1}^{U_i} c\right) + U_i}{\frac{F_i}{2}} \tag{7.10}$$

The first two terms correspond to the cost of generating all the time-delay blocks of all the outputs in the L bands. The third term shows the processing cost of the synthesis bank to generate the full-band output blocks. The fourth term is the cost of obtaining the correlation submatrices $R_{mn}^{i}(l)$ (see equation (7.7)) for calculating the gradient. The fifth term indicates the cost of adaptation considering only the updating of the U_i coefficients of the separation filter in the various bands. The denominator factor of all plots indicates that operations are being conducted on a sample rate $\frac{F_i}{2}$ times lower than the rate at full-band. For separation filters of higher orders, the above equation can be simplified by taking the effective parts of equation (7.7), it is written as

$$N_{blockMB} \approx M^2\sum_{i=1}^{L}\left(\frac{8U_i^3}{F_i}\right). \tag{7.11}$$

The above results were obtained using the simplified normalisation factor (equation (7.9)), $D_i = U_i$ and $K_i = 2\,D_i$ to the multiband algorithm.

7.2.2 *Experimental Results*

Through computer simulations, the performance of our NMD_UEPS filter bank is evaluated and its validity is verified for BSS.

Two speech signals are used in all experiments with sampling frequency f_s = 8kHz and 10 seconds duration , two voices, one female and the other male, are uttered in English. These speech signals were convolved with impulse responses [**77**] in a virtual room scenario that is discussed in Section 2.2.1.2.1.

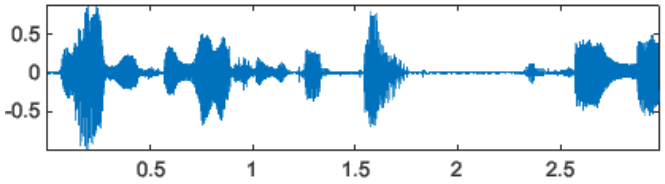
In our experiments, the mixtures were performed considering different reverberation conditions. The impulse responses were convolved with the signals as in [**152**] for f_s = 8kHz. Then, they are truncated to the desired size. The mixing filters' length ($L = U$) is fixed. The algorithms are run on Intel i3 processor with 2.53 GHz speed. For performance evaluation, we have used the signal-to-interference ratio (SIR) defined in Section 5.3.

The separation filter coefficients of each band $\widetilde{\omega}_{mn}^{i}(k)$ were always initialised to zero except for $m = n$, and $k = 0$, which were made equal to 1. The same procedure is adopted for the full-band algorithm. Fig. 7.2 shows the original sources, mixtures, and estimated waveforms.

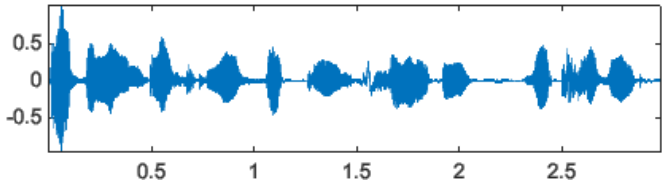
7.2.2.1 ***Experiment 1***

In this experiment, a comparison is performed in full-band between (a) a novel normalisation (equation (7.1)) and (b) an old normalisation (equation (6.53)) [**53**][12]. Fig. 7.3, shows the performance of the SIR considering mixture lengths L = 64, 128 and 256. The rate is $\gamma = 5 \times 10^{-3}$, but for the new normalisation L is 256. For convergence to be reached we use γ to be 0.001. Table 7.1 presents the simulation time for that shown in Fig. 7.3. Analysing the results of Table 7.1 and Fig. 7.3, we can see that for longer length mixing filters (corresponding to more reverberating environments), the new normalisation scheme substantially enhances the speed of convergence of the full-band approach and significantly reduces the simulation time.

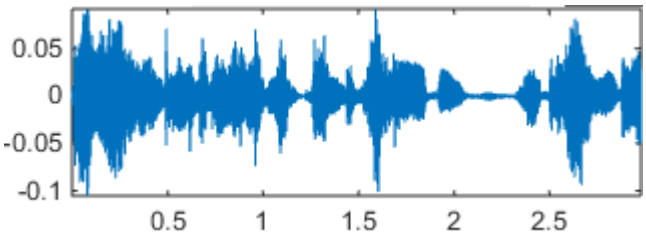
[12] We used this normalisation to compare the results as our work is an extension from [**53**].



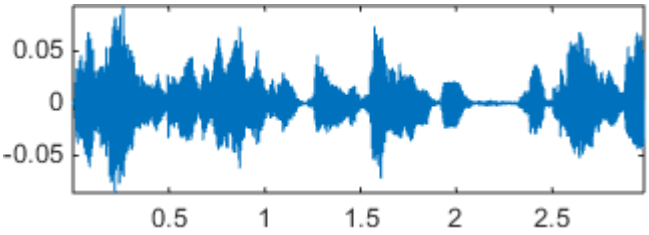
(a)



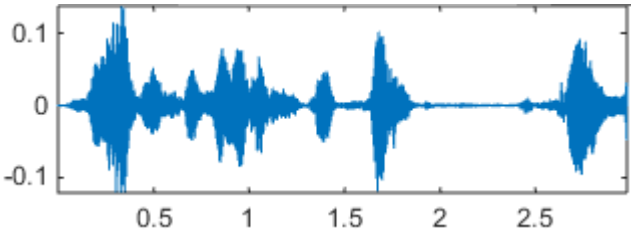
(b)



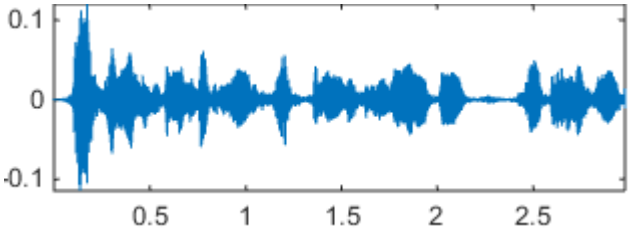
(c)



(d)

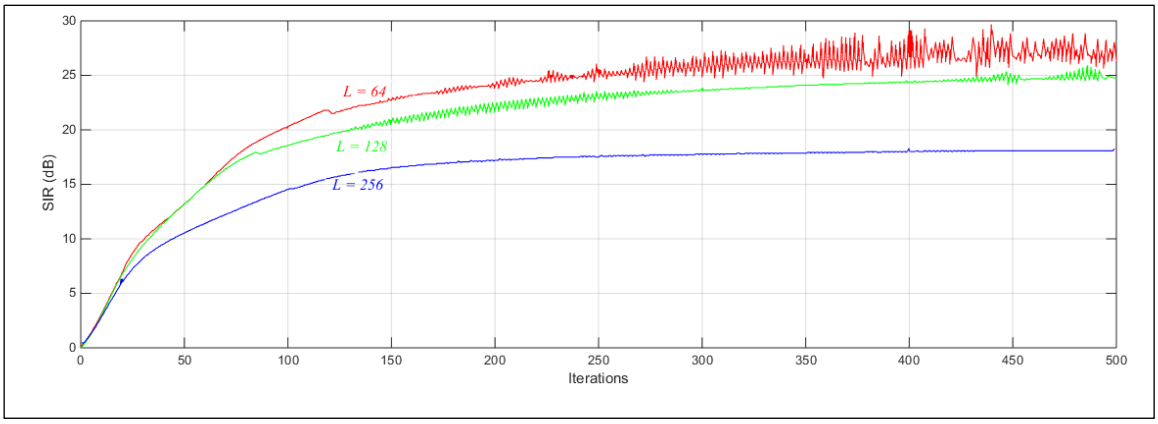


(e)

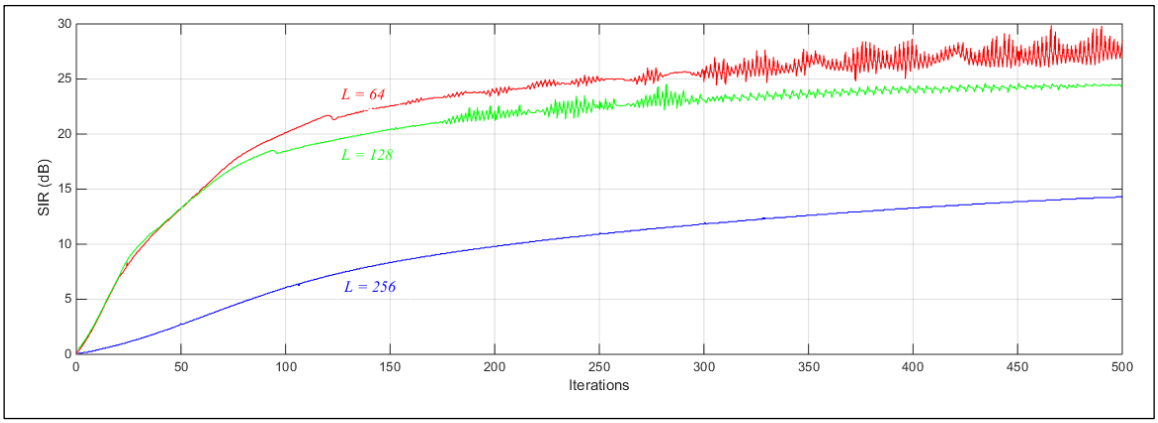


(f)

Fig. 7.2 (a) source 1, (b) source 2, (c) mixture 1, (d) mixture 2, (e) estimated source 1, (f) estimated source 2



(a)



(b)

Fig. 7.3: Comparison between the two normalisations: (a) the new and (b) the old one.

Table 7.1 Simulation time in minutes

U = *L*	**New Normalisation**	**Old Normalisation**
64	11	19
128	27	63
256	64	178

7.2.2.2 ***Experiment 2***

In this experiment, discarding the use of a filter bank with complex coefficients, we contrast only the full-band and multiband schemes, both with approximate normalisation methods according to equations (7.1) and (7.9). The UEPS is tested with octave-band filter banks and perfect reconstruction for L = 4 bands.

Fig. 7.4 shows the frequency responses of the analysis filter $P_{i,j}$(z) of the multiband UEPS, shown in Fig. 7.1. Fig. 7.4 shows the frequency responses of the analysis filters Pi,$_{i(}$z*)* of the multiband UEPS shown in Fig. 7.1.

Table 7.2 shows the decimation factors $\mathcal{F}_i$, the orders of analysis filters K_{P_i} (which are of equal orders of synthesis filters K_{Q_i}), the delays d_i, the orders U_i of the separation filter ω_{mn}^i and learning rates γ_1^i and γ_2^i, used in multiband simulations with L = U = 1024. The learning rate of the algorithm for full-band was the same as in Experiment 1, except for (L = 1024) where ($\gamma = 3 \times 10^{-3}$) was used. These steps resulted in better convergence and are obtained experimentally by trial and error. Fig. 7.4 shows the frequency responses of the analysis filters $P_{i,i}(z)$ of the multiband UEPS shown in Fig. 7.1.

Table 7.2 Parameters of the scheme with UEPS
for L = 4 and U = 1024

i	F_i	K_{P_i}	d_i	U_i	γ_1^i	γ_2^i
0	8	381	0	337	0.0092	0.009
1	8	381	0	337	0.0159	0.019
2	4	177	243	700	0.0377	0.037
3	2	72	365	603	0.0714	0.071

To minimise the complexity without devolution in the separation stage, the order of the separation filter in the higher frequency band (i = 3), which is adjusted to the same

rate as the signals captured by the microphones, is reduced and compared to that of equation (7.5) for[13]

$$U_3 = 2\left[\frac{\frac{U-1}{2} + K_{Q_i}}{F_i}\right]. \tag{7.12}$$

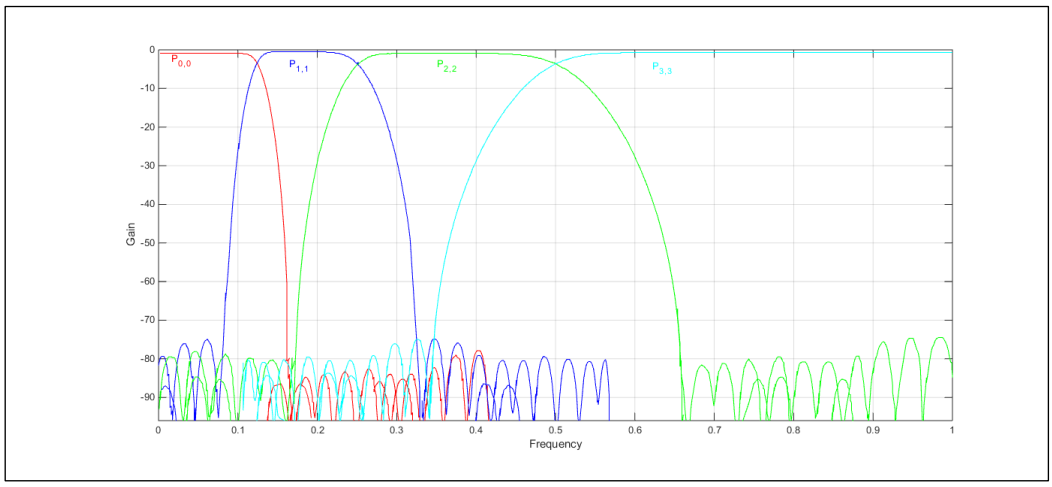
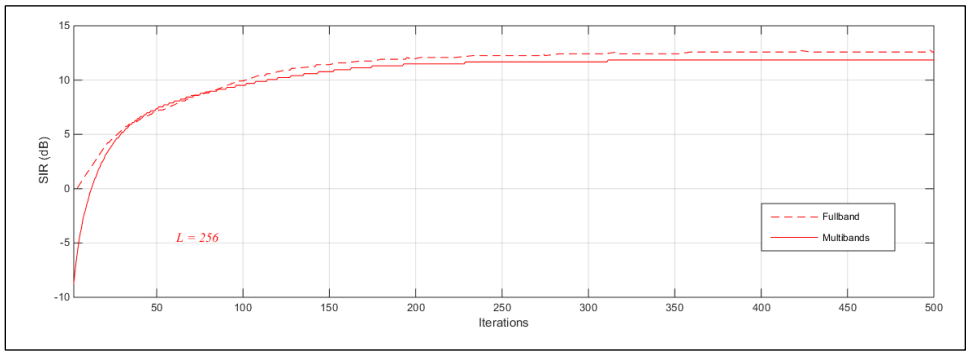


Fig. 7.4: The frequency responses of the analysis filters $P_{i,i}(z)$ for $L = 4$.

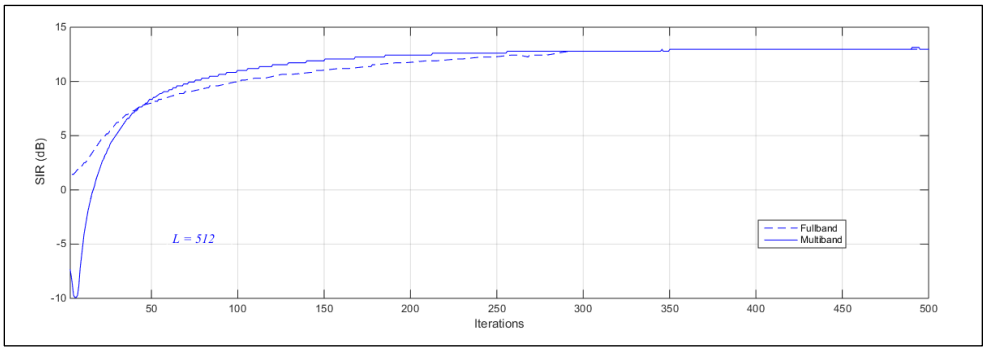
The above result was obtained by trial and error. The worst band in multiband BSS was band number 3. There was a reduced signal-to-interference ratio caused by the use of short length separation filters. This reduction was possibly due to the reverberation characteristics and the small concentration of speech energy signals at high frequencies.

Mixing filters of lengths L = 256, 512 and 1024 are used for the full-band and multiband algorithms in the equations (6.52) and (7.6), respectively, and compared in Fig. 7.5 for the SIR performance.

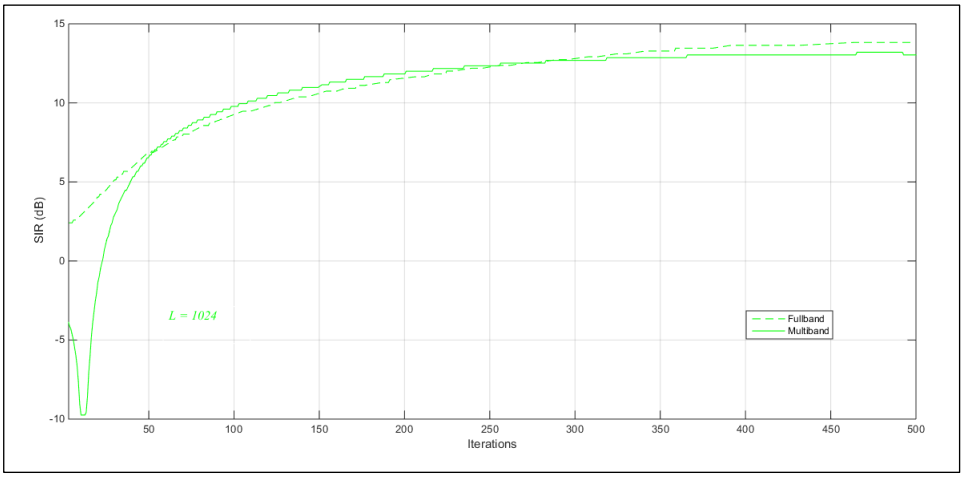
[13] It is known that the decay of the impulse responses of rooms is faster at high frequencies. Therefore, we observed that we could consider that the unknown system had a reduced size at the high-frequency band. In this way, we saved in computational complexity. This is the reason for the change from over equation (7.12) to equation (7.5).



(a)



(b)



(c)

Fig. 7.5: The effect of Signal-to-Interference Ratio of FB and multiband Schemes with filters of different *L*s

Table 7.3 shows the best signal-to-interference-ratio obtained for full-band and four bands of the UEPS.

Table 7.4 presents the simulation time that is depicted in Fig. 7.5, and the computational cost of the full-band (FB) multiband (MB) algorithms in accordance with equations (7.2) and (7.11), respectively. We can conclude from these tables and the results in Fig. 7.5 that at the higher mixing system order (higher reverberant environments), the better multiband scheme's performance is achieved in contrast to the full-band scheme; thus, better rate[14] of convergence is obtained with lower computational cost and a shorter simulation time.

Table 7.3 Signal-to-Interference-Ratio

Mixtures	**Signal-to-Interference (dB) for different bands** k **=**				**SIR (dB)**
U = S	**0**	**1**	**2**	**3**	
256	20.69	12.11	13.07	6.81	14.77
512	13.91	8.55	12.19	6.81	10.57
1024	13.60	8.98	10.09	6.34	9.99

Table 7.4: Simulation time in minutes

Mixing Filters	**Full-band**		**MultiBand**	
U = S	Simulation time	N_{block}	Simulation time	N_{block}
256	55	2.85×10^8	53	4.90×10^8
512	137	2.01×10^9	97	1.96×10^9
1024	389	1.83×10^{10}	176	1.01×10^{10}

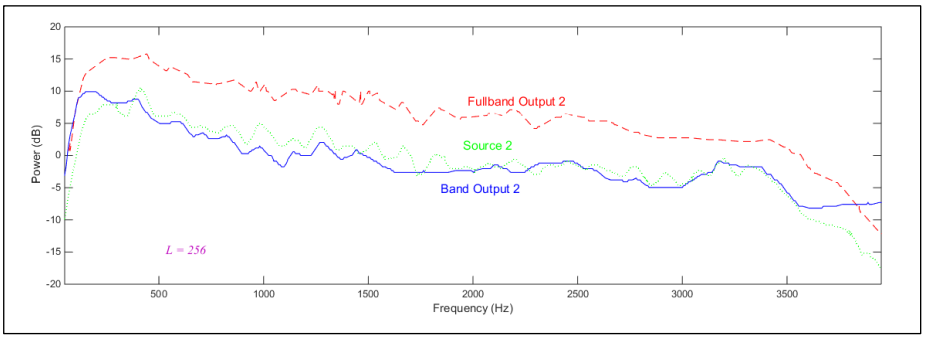
[14] The convergence rate is quantified visually, through SIR performance; we must take into account the final SIR value (steady-state) and the amount of iterations (iteration number) to get this result. In other words, once SIR is no longer improving, or is growing very slowly, the convergence is almost complete.

Having done these tests for multiband scheme over the full-band scheme, now we will validate our proposed algorithm with some standards and another best performing work using PESQ metrics as a benchmark.

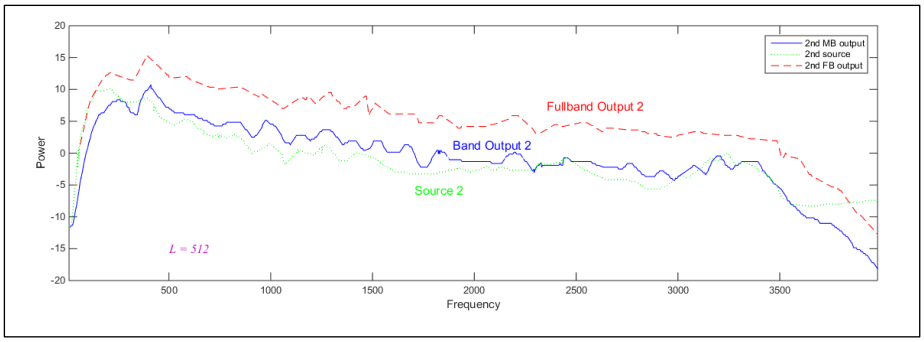
The performance of the BSS in $i = 3$ is always inferior to the other bands, because of employing a reduced filter's length, leading to a minimised signal-to-interference-ratio. The perceptual evolutions are carried out by one female and one male English speakers using a PESQ (see Section 2.7.5) tool [**153**] and the results are compared with three standards, that are used as a benchmark, and one best performing work [**81**], see Table 7.5. Fig. 7.6 depicts the estimates of the original sources of full-band and multiband for following lengths of mixtures: L = 256, 512, and 1024, and their spectral power. Fig. 7.7 provides visualization for the spectrum of the first and second original sources compared to the spectrum of the estimated signals at the outputs of the first and second band, respectively. The Matlab code of this visualization is given in Appendix A.4. The obtained outcomes demonstrate how resilient the proposed scheme is to the scaling of the output signals and the whitening of the sources.

Table 7.5: The PESQ of the proposed method is compared to three standards and best performing work [**81**].

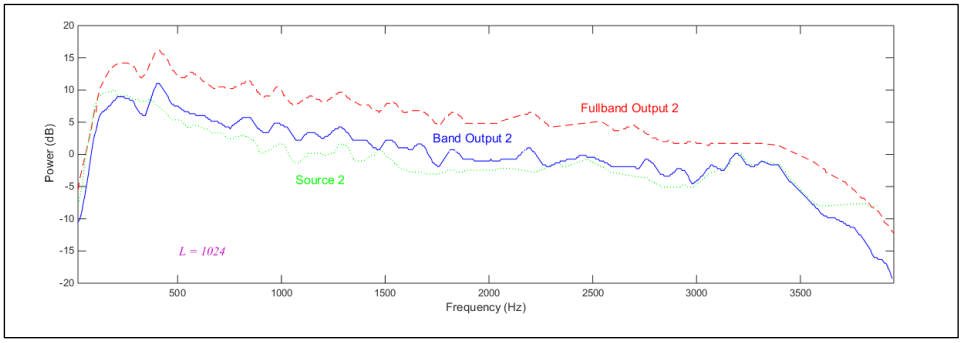
	FastICA	**JADE**	**SOBI**	[**81**]	**Proposed Method**
PESQ of the Female Speaker	3.25	3.29	2.58	3.29	3.31
PESQ of the Male Speaker	4.27	4.14	3.45	4.38	4.43



(a) Mixing filter length 256



(b) Mixing filter length 512



(c) Mixing filter length 1024

Fig. 7.6: FB and MBs spectrum with different filter lengths

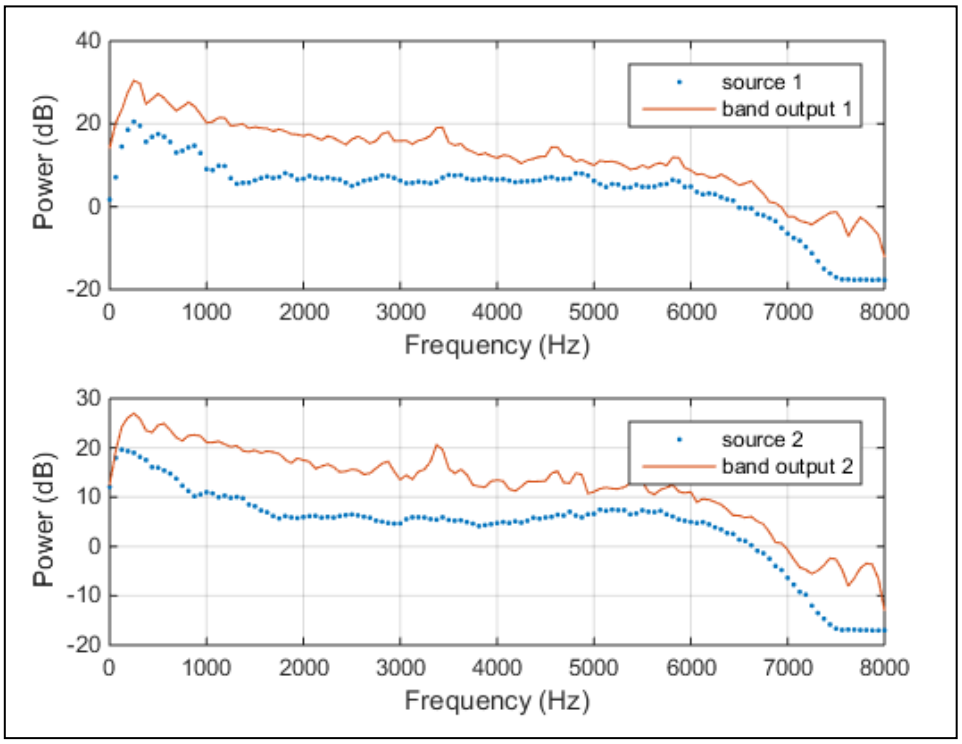


Fig. 7.7 The spectra of the original sources 1 and 2, and the separated signals at bands y1 & y2.

7.3 Conclusions

This chapter discussed the proposed multiband algorithm for BSS, employing UEPS filter banks. In this algorithm, various separation filters' lengths are used and are adapted at various rates.

Simulations are performed on speech signals. These simulations demonstrate the superior performance of the proposed multiband scheme over the full-band scheme. The proof was made in terms of the simulation time, the convergence rate and the SIR ratio.

Chapter 8
Concluding Remarks, Limitations and Future Work

In this chapter, we present the concluding remarks of this work as well as the limitations to the proposed algorithm. Finally, we conclude with some suggestions for future work. The main focus of this work is to obtain an UEPS with multiple bands. The goal is to improve the convergence rate and the computational complexity. This chapter is divided into three subsections: concluding remarks, limitations, and future work.

8.1 Concluding Remarks

The experimental results involving audio signals showed an improvement in signal interference and a reduction in the computational cost obtained from the multiband schemes. Some remarks are presented in the following two subsections:

8.1.1 *A MD_UEPS*

An adaptive scheme of a MD_UEPS is derived:

- Such a scheme helps in breaking down the signal at the input. Breaking down using multiple bands helps to reduce the sampling rate. Reducing the sampling rate helps to improve the convergence rate.
- An LMS method is proposed that uses a normalised adaptation-step to deal with the reduction in the sampling rate. This LMS method is used to minimise the errors in the bands and update the coefficients of the proposed scheme. This method works at a low sampling rate between the bands and uses a normalised learning

rate. A formula of best selected taps is derived; these taps are selected from the proposed scheme in order to model an FIR system.

- Analysis of the convergence is done; we concluded that the convergence rate of the proposed method is controlled by a matrix that contains the coefficients of the analysis filters. This helps us to determine the nature of the proper analysis bank and define the correct number of bands by observing the convergence rate.
- Mean square error is analysed; it is analysed in a stationary state of the proposed method by taking into consideration the error raised by the non-overlapping between the non-adjacent bands of the analysis filters. However, the remaining errors are modelled by the conventional analysis of the LMS.
- Four experiments were conducted:
    - The first experiment was to compare various bands of the proposed scheme and observe the behaviour of the mean square error for EPS and UEPS. This experiment showed that the use of multiple bands with the proposed UEPS makes the convergence faster than that of the EPS. The reason for the improvement in the convergence speed is the use of multiple bands at the input. The signal at the input is broken down among the bands causing a decrease in the power density of the signals, and subsequently, an enhancement in the convergence speed.
    - In the second experiment, a white noise is applied at the input and the convergence speed is tested with two distinct bands of the proposed scheme. This experiment demonstrates how resilient the proposed scheme is to white noise.
    - The third experiment tested the proposed algorithm against fast changes to the unknown system by adding variance to

the white noise of the required signal. The results showed that the proposed algorithm is robust in detecting quick changes in the system and has a similar performance to the EPS.

- The fourth experiment is done to study the theoretical and experimental mean square errors in the system identification. The theoretical and practical results were very similar to each other.

8.1.2 *A NMD_UEPS*

- An oversampled UEPS is proposed. This aims to improve the convergence rate and decrease the computational cost of the schemes that need very large adaptive taps, such as those that are used in audio applications, for example acoustic echo cancellation. Two experiments are carried out:
  - A comparison is made, in the first experiment, between the old and the proposed scheme using different measures:
    - The simulation time that is observed for different filter lengths,
    - The signal-to-interface ratio for different filter lengths.

The results showed the proposed scheme is faster than the old one.

- A comparison is made, in the second experiment between the full-band and multiband approaches: adaptation-sizes are measured for the quickest convergence,
- Reduced computational cost is achieved by reducing the orders of the separation filter without corrupting the separation process.

8.2 Limitations

There are some limitations to the proposed contributions:

- The number of speakers that are used in our proposed scheme is limited to two; increasing the number of speakers will make the mathematical equations more complicated and hence the experimental results will be unachievable. The reasons are:
    - Overlapping between bands will be unavoidable.
    - The mean square errors will be increase and the adaptive filters will not be able to optimise the errors.
    - The Increased number of taps.
    - The hardware circuit will be too big and impractical.
    - The complexity cost will be higher.
    - Cannot reach a feasible signal-to-interference ratio.
    - The convergence speed cannot be achieved.
- Reverberation environments limit the use of the proposed algorithms, as the results are obtained for a specific room with specific dimensions.
- Microphones being placed too far apart may affect the results.
- The results may change if the experiments were conducted outdoors.
- The proposed algorithms may be unreliable for other languages.

8.3 Future Work

Based on our work, the following studies could be carried forward:

1. Explore efficient ways of implementation for cosine modulated filter bank;
2. Develop prototypes which minimise the final mean square error of the scheme and/or accelerate the convergence rate;
3. To evaluate the behaviour of multiband schemes in acoustic echo cancellation using dedicated prototypes;
4. Derive a maximally-decimated UEPS to be used in BSS applications;

5. Investigate the use of higher order statistics in the objective function that is used in blind source separation;
6. Test the use of conventional wavelet functions in UEPS;
7. Experiments can be conducted with languages other than English.
8. Experiments can be performed for different genders and different ages.
9. Experiments can be conducted in stressful and emotional talking environments.

Appendix A

Appendix A.1

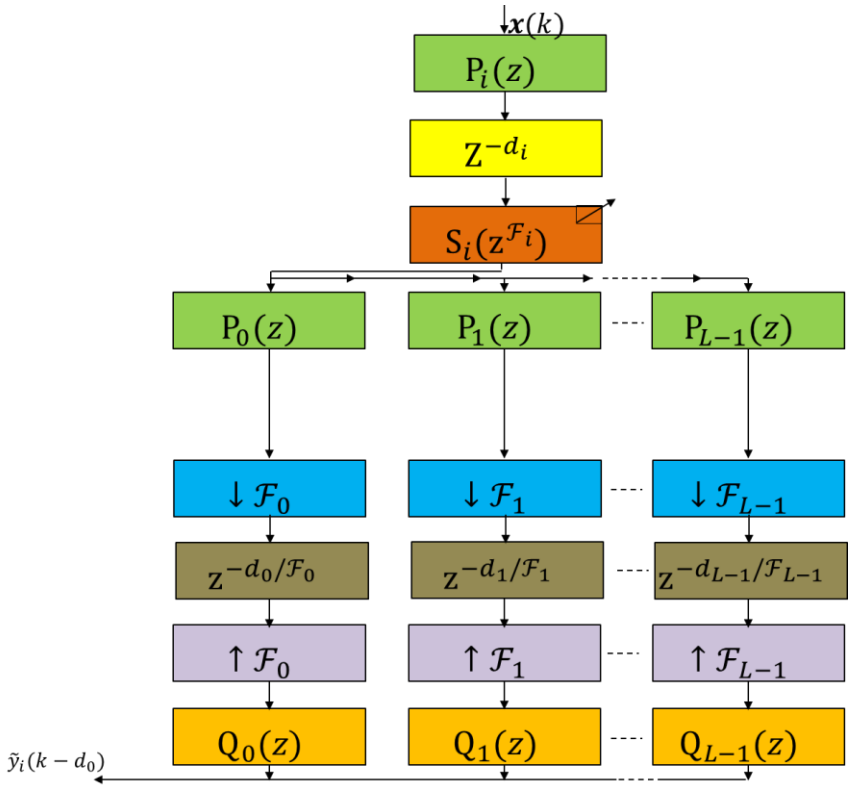
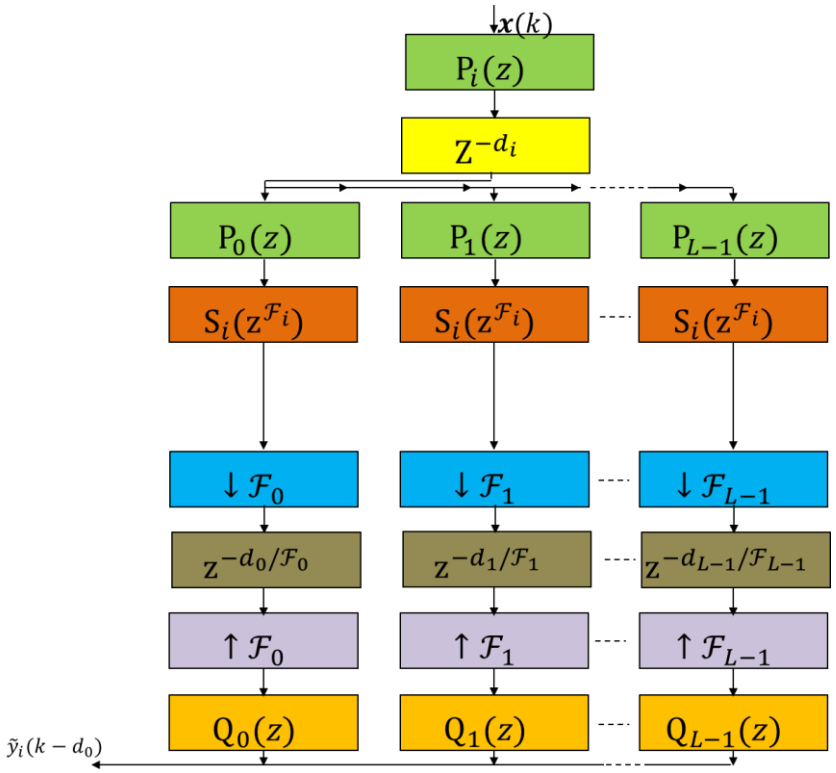
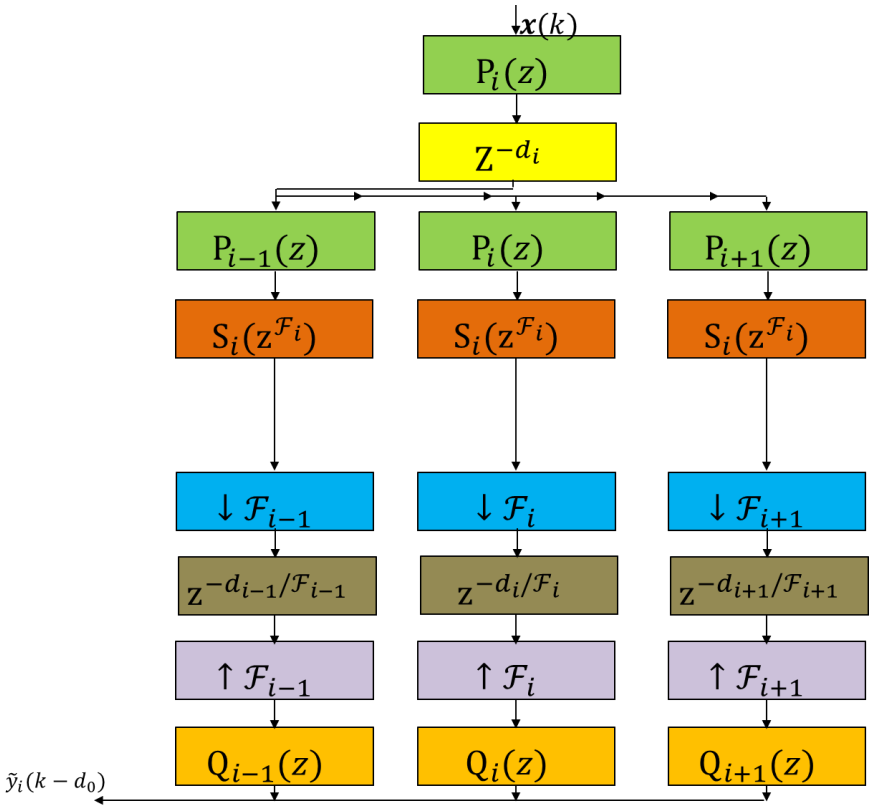


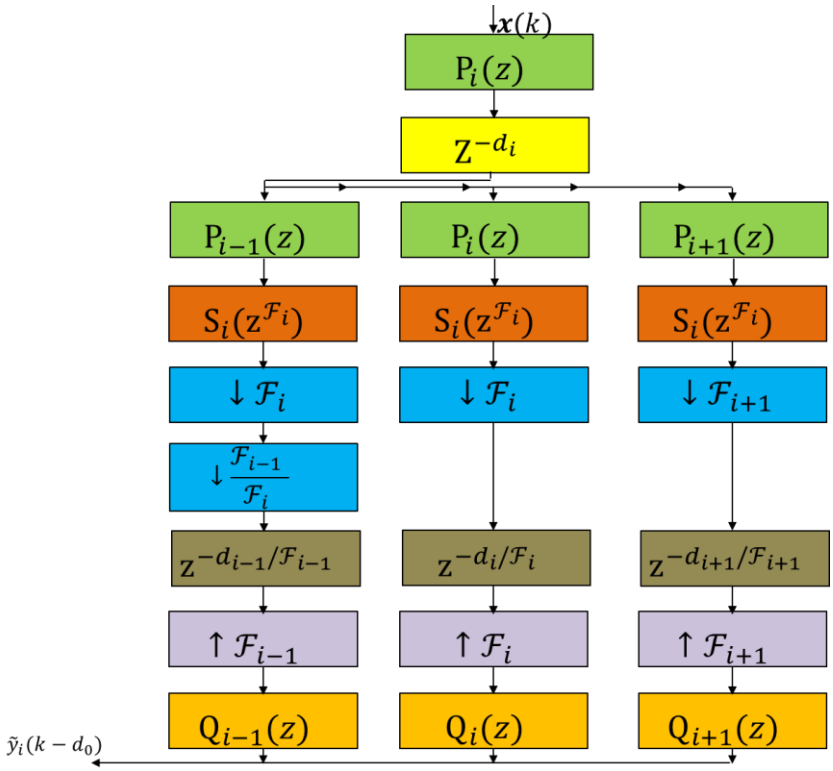
Fig. 5.2 Incorporating maximally decimated filter bank of the i^{th} band.



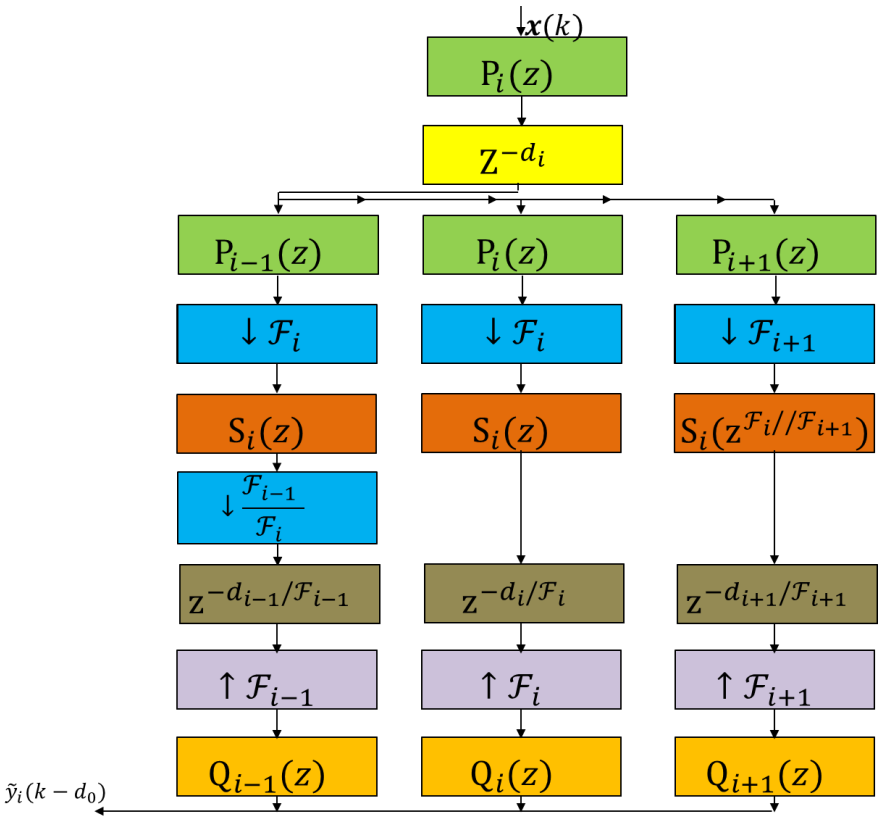
Step 1: using assumption A3, changing of the orders of S_i and P_j in Fig. 5.2.



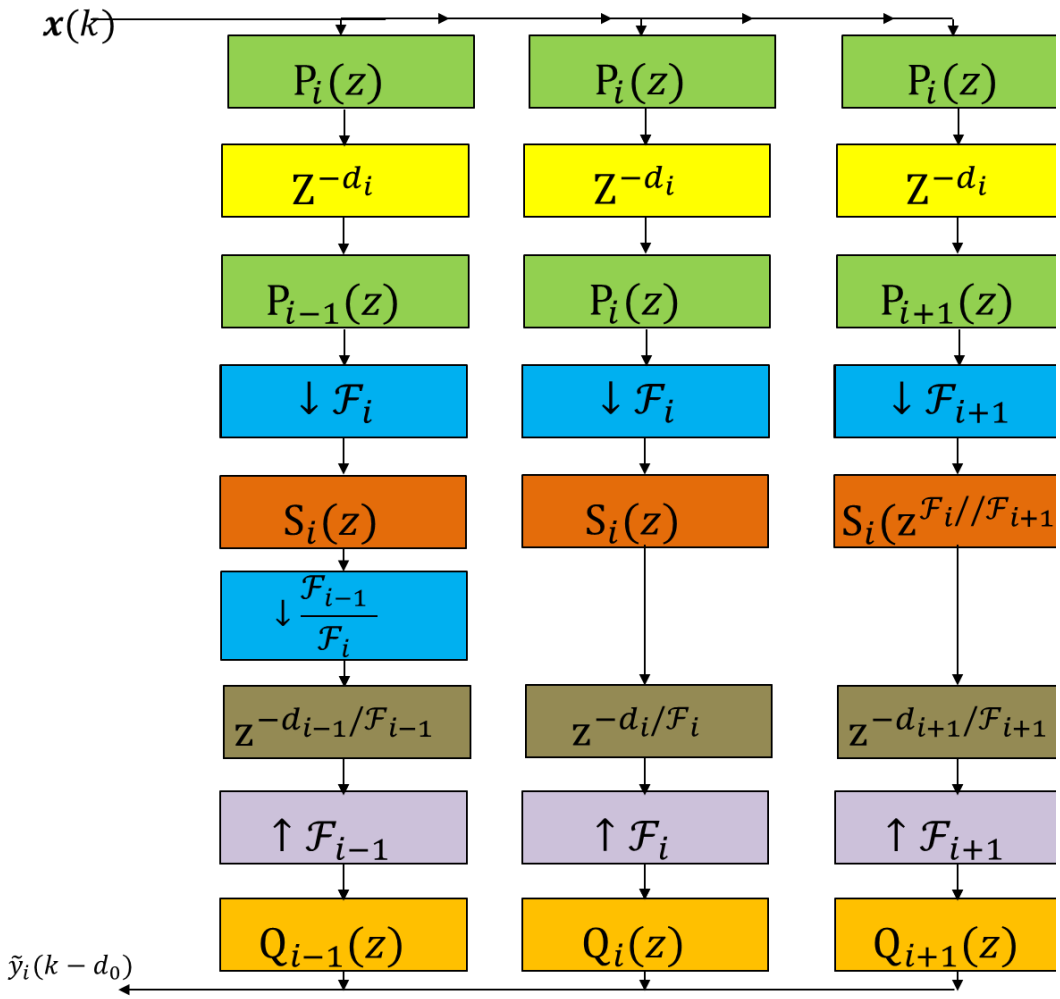
Step 2: using assumption A1, keep only the paths corresponding to adjacent analysis filters, that is, only $P_i(z)$ followed by $P_{i-1}(z)$, $P_k(z)$ and $P_{i-1}(z)$.



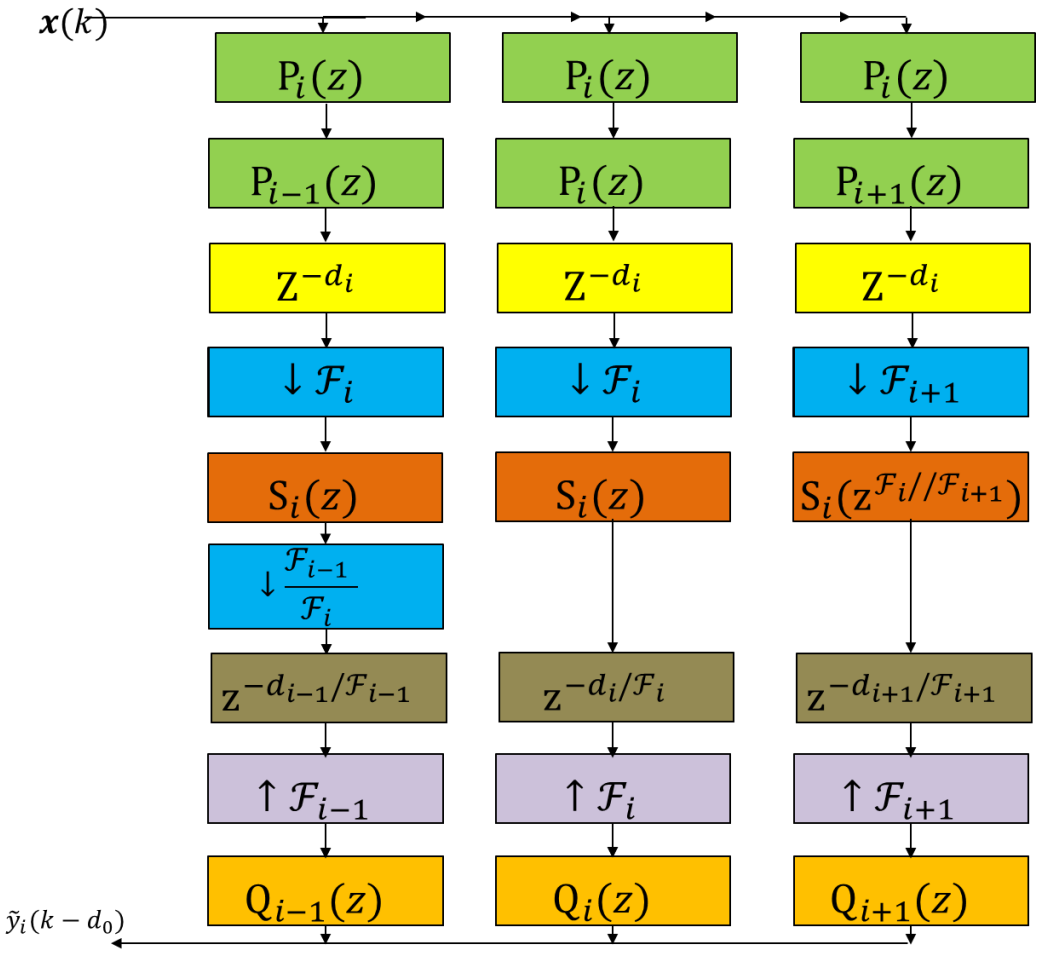
Step 3: from assumption A2, the decimation by $\mathcal{F}_{i-1}$ can be implemented as decimation by $\mathcal{F}_i$ followed by decimation by $\frac{\mathcal{F}_{i-1}}{\mathcal{F}_i}$.



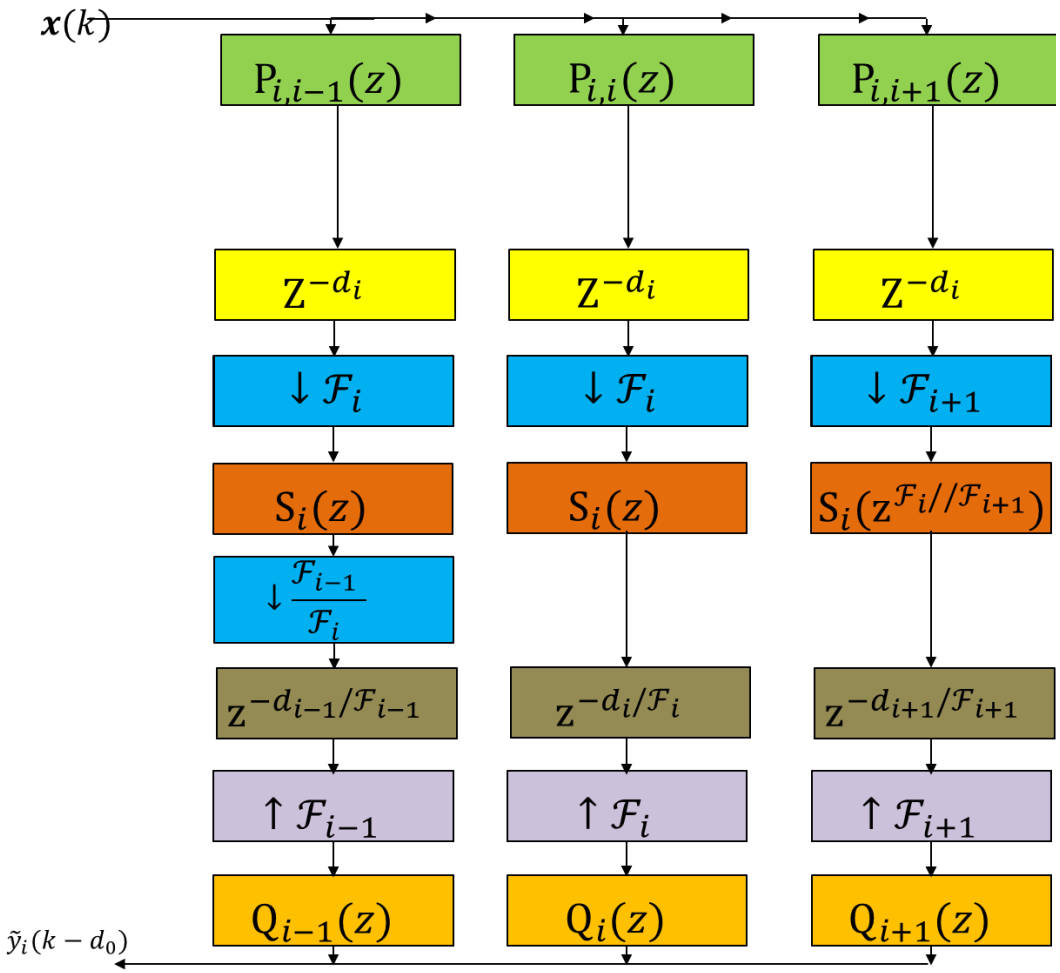
Step 4: using the first Noble Identity, move $S_i(z^{\mathcal{F}_i})$ to the above of the decimators $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$.



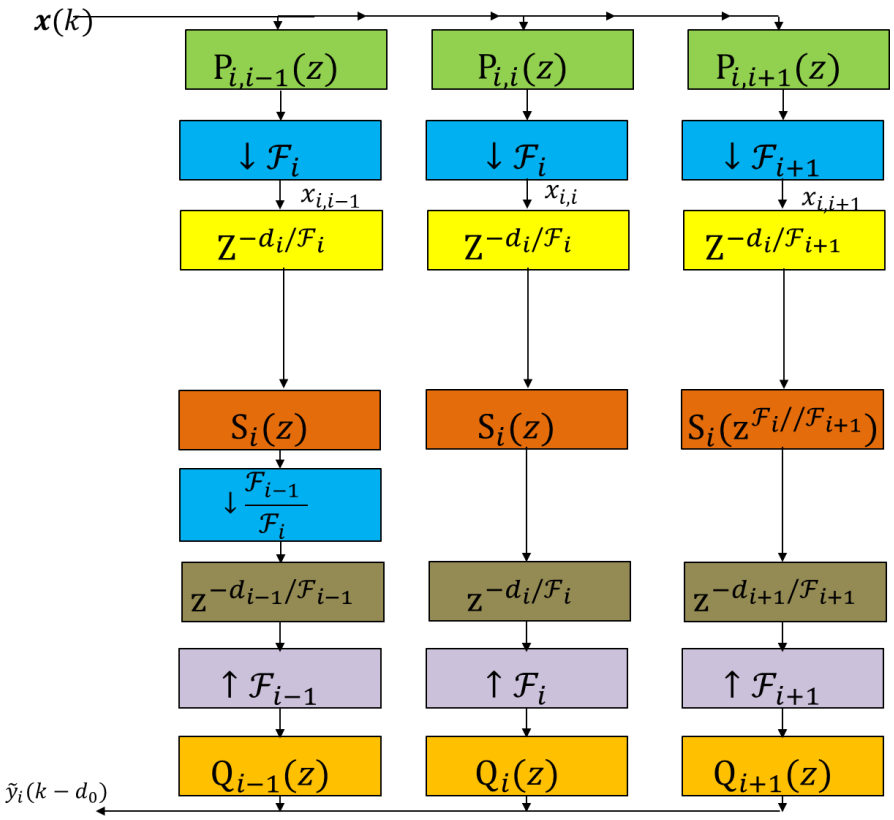
Step 5: The change in the order of the delays and filters is allowed, since there is no sampling rate modification, and, therefore, the delays and filters are time invariant systems.



Step 5 is implemented as shown in the figure above.



Step 6: rearranging the subscripts of *P*.



Step 6: Again using Nobel Identity we obtain Fig. 5,3.

Appendix A.2

i	r	$\begin{cases} F_0 = 2^{L-1} \\ F_i = 2^{L-i} \quad for\ 1 \le i < L-1 \end{cases}$	$\ell_i = \frac{\mathcal{F}_0}{\mathcal{F}_i}$	$\tilde{S}_{\ell_i+r}(z)$	$S_{i,r}(z)$
1	0	$\mathcal{F}_1 = 8$	$\ell_1 = 1$	$\tilde{S}_1(z)$	$S_{1,0}(z)$
2	0	$\mathcal{F}_2 = 4$	$\ell_2 = 2$	$\tilde{S}_2(z)$	$S_{2,0}(z)$
2	1	$\mathcal{F}_2 = 4$	$\ell_2 = 2$	$\tilde{S}_3(z)$	$S_{2,1}(z)$
3	0	$\mathcal{F}_3 = 2$	$\ell_3 = 4$	$\tilde{S}_4(z)$	$S_{3,0}(z)$
3	1	$\mathcal{F}_3 = 2$	$\ell_3 = 4$	$\tilde{S}_5(z)$	$S_{3,1}(z)$
3	2	$\mathcal{F}_3 = 2$	$\ell_3 = 4$	$\tilde{S}_6(z)$	$S_{3,2}(z)$
3	3	$\mathcal{F}_3 = 2$	$\ell_3 = 4$	$\tilde{S}_7(z)$	$S_{3,3}(z)$

Appendix A.3

To clarify the results obtained in Section 0, we consider an unknown system identification with the proposed scheme assuming a decomposition of $L = 3$ bands. According to equations (5.4) and (5.5), we can write the relationship between the multiband coefficients and the unknown system (see equation (5.11) assuming $d_m = 0$) as follows:

$$\begin{bmatrix} S_{0,0}(z) \\ S_{1,0}(z) \\ S_{2,0}(z) \\ S_{2,1}(z) \end{bmatrix}^T = \begin{bmatrix} U_0(z) \\ U_1(z) \\ U_2(z) \\ U_3(z) \end{bmatrix}^T \begin{bmatrix} Q_{0,0}(z) & U_{1,0}(z) & U_{2,0}(z) & U_{3,0}(z) \\ U_{0,1}(z) & U_{1,1}(z) & U_{2,1}(z) & U_{3,1}(z) \\ U_{0,2}(z) & U_{1,2}(z) & U_{2,2}(z) & U_{3,1}(z) \\ U_{0,3}(z) & U_{1,3}(z) & U_{2,3}(z) & U_{3,1}(z) \end{bmatrix} \tag{A.3.1}$$

Then, from equation (A.3.1), the optimal coefficients can be found as given by equation (5.13), i.e.:

$$\begin{cases} S_0(z) = S_{0,0}(z) \\ S_1(z) = S_{1,0}(z) \\ S_2(z) = S_{2,0}(z^2) + z^{-1} S_{2,1}(z^2), \end{cases} \tag{A.3.2}$$

where $\mathcal{F}_0 = \mathcal{F}_1 = 4$, $\mathcal{F}_2 = 2$, $\ell_0 = \ell_1 = 1$ and $\ell_2 = 2$.

Appendix A.4

```
% Spectrum of the original sources (s1 & s2)
 [BS1,FF,T]=specgram(s1,[],FS);
 [BS2,~,~]=specgram(s2,[],FS);
% Spectrum of the estimated sources (y1 & y2)
[By1,~,~]=specgram(y1,[],FS);
[By2,~,~]=specgram(y2,[],FS);
for f=1:length(FF)
    PS1(f)=norm(BS1(f,:));
    PS2(f)=norm(BS2(f,:));
    PY1(f)=norm(By1(f,:));
    PY2(f)=norm(By2(f,:));
end
figure, subplot(2,1,1),plot(FF,10*log10(PS1),'.',FF,10*log10(PY1))
xlabel('Frequency (Hz)'), ylabel('Power (dB)'), legend('source 1','band
output 1'), grid
subplot(2,1,2),plot(FF,10*log10(PS2),'.',FF,10*log10(PY2))
xlabel('Frequency (Hz)'), ylabel('Power (dB)'), legend('source 2','band
output 2'), grid
```

References

[1] Paulo Diniz, *Adaptive Filtering Algorithms and Practical Implementation*, 3rd ed. New York, United States: Springer, 2008.

[2] Sebastian Ewert and Bryan Pardo, "Score-Informed Source Separation for Musical Audio Recordings: An overview," *Signal Processing Magazine, IEEE*, vol. 31, no. 3, pp. 116 - 124, October 2014.

[3] Wing-Kin Ma and Jose Bioucas-Dias, "A Signal Processing Perspective on Hyperspectral Unmixing: Insights from Remote Sensing," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 67 - 81, January 2014.

[4] Aishwarya Moni, Ivan Lokmer, Christopher Bean, and Scott Rickard, "Source Separation on Seismic Data: Application in a Geophysical Setting," *Signal Processing Magazine, IEEE*, vol. 29, no. 3, pp. 16 - 28, May 2012.

[5] Rajib Jha and Prabir Biswas, "Contrast Enhancement of Dark Images Using Stochastic Resonance," *Image Processing, IET*, vol. 6, no. 3, pp. 230 - 237, April 2012.

[6] Chen Wang, Shenhua Huang, Xuchu Dai , and Peixia Xu, "A Deterministic Blind Separation Approach Based on GA for Digital Communication Signals," *IEEE International Symposium on Communications and Information Technology, ISCIT*, vol. 1, no. 1, pp. 495 - 498, November 2005.

[7] Udo Zölzer, *Digital Audio Signal Processing*, 2nd ed. Sussex, UK: Wiley, 2008.

[8] Choong-Hwan Chio and Won Chang, "Blind Source Separation of Speech and Music Signals Using Harmonic Frequency Dependent Independent Vector Analysis," *IEEE Electronics Letters*, vol. 48, no. 2, pp. 124 - 125, January 2012.

Jorge Marin-Hurtado and Devangi Parikh, "Perceptually Inspired Noise-

9] Reduction Method for Binaural Hearing Aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1372 - 1382, May 2012.

Jeong-Sik Park and Gil-Jin Jang, "Acoustic Interference Cancellation for a
10] Voice-Driven Interface in Smart TVs," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 244 - 249, April 2013.

David Peterson, James Knight, and Michael Kirby, "Feature Selection and
11] Blind Source Separation in an EEG-Based Brain-Computer Interface," *EURASIP Journal on Applied Signal Processing*, vol. 19, no. 1, pp. 3128–3140, December 2005.

Longbiao Wang, Kyohei Odani, and Atsuhiko Kai, "Speech Recognition
12] Using Blind Source Separation and Dereverberation Method for Mixed Sound of Speech And Music," in *Signal and Information Processing Association Annual Summit and Conference*, Kaohsiung, 2013, pp. 1 - 4.

Simon Haykin and Zhe Chen, "The Cocktail Party Problem," *Neural
13] Computation*, vol. 17, no. 9, pp. 1875-1902, September 2005.

Simon Haykin, *Adaptive Filter Theory*, 4th ed. New Jersey, United States:
14] Prentice Hall, 2002.

Pierre Comon, "Independent Component Analysis, a New Concept," *Signal
15] Processing, Elsevier*, vol. 36, no. 3, pp. 287 - 314, March 1994.

Kostas Kokkinakis and Asoke Nandi, "Multichannel Blind Deconvolution
16] for Source Separation in Convolutive Mixtures of Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 200 - 212, March 2006.

Michael Pedersen, Jan Larsen, Ulrik Kjems, and Lucas Parra, *A Survey of
17] Convolutive Blind Source Separation Methods*, 1st ed. New York, United States: Springer Press, 2007.

Kari Torkkola, "Blind Separation of Convolved Sources Based on

18] Information Maximization," in *IEEE Workshop on Neural Networks and Signal Processing*, Kyota, 1996.

Anthony Bell and Terrence Sejnowski , "An Information-Maximization
19] Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129-59, April 1995.

Te-Won Lee, "Blind Separation of Delayed and Convolved Sources," in
20] *Proceedings Advances in Neural Information Processing Systems*, Cambridge, MA, 1997, pp. 758-764.

Ryo Mukai , Shoko Araki, and Shoji Makino, "Separation and
21] Dereverberation Performance of Frequency Domain Blind Source Separation," in *Proceedings Independant Component Analysis (ICA'01)*, Kyoto, 2001, pp. 230-235.

Ryo Mukai, Shoko Araki, Hiroshi Sawada, and Shoji Makino, "Evaluation
22] of Separation and Dereverberation Performance in Frequency Domain Blind Source Separation," *Journal of Acoustic Science and Technology*, vol. 25, no. 2, pp. 119-126, November 2004.

Paris Smaragdis, "Blind Separation of Convolved Mixtures in the
23] Frequency Domain," *Neurocomputing*, vol. 22, no. 1, pp. 21-34, November 1998.

Siavash Ivrigh, Seyed Sadough , and Seyed Ghorashi, "A Blind Source
24] Separation-Based Positioning Algorithm for Cognitive Radio Systems," *Journal of Applied Sciences Engineering and Technology* , vol. 4, no. 4, pp. 299-305, February 2012.

Sergio Netto and Paulo Diniz, *Digital Signal Processing*, 2nd ed. New
25] York, United States: Cambridge University Press, 2010.

Seong-Eun Kim, Young-Seok Choi, and Moon-Kyu Song, "A Subband
26] Adaptive Filtering Algorithm Employing Dynamic Selection of Subband Filters," *Signal Processing Letters, IEEE*, vol. 17, no. 3, pp. 245 - 248, March 2011.

Yong Lian and Ying Wei , "A computationally Efficient Nonuniform FIR
27] Digital Filter Bank," *Computers & Digital Techniques, IET*, vol. 2, no. 4, pp. 285 - 294, December 2009.

Julius Smith, "AUDIO FFT Filter Banks," in *Digital Audio Effects*, Como,
28] 2009, pp. 11-18.

Bruno Paillard, Joel Soumagne, Philippe Mabilleau, and Sarto Morissette,
29] "Subband Decomposition: An LMS-Based Algorithm to Approximate The Perfect Reconstruction Bank in The General Case," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 233 - 238, January 1991.

André Gilloire and Martin Vetterli, "Adaptive Filtering in Subbands with
30] Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancellation," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 1862-1875, August 1992.

Yoji Yamada, Hiroshi Ochi, and Hitoshi Kiya, "A Subband Adaptive Filter
31] Allowing Maximally Decimation," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 9, pp. 1548-1552, December 1994.

Behrouz Farhang-Boroujeny and Zhongjun Wang, "Adaptive Filtering in
32] Subbands: Design Issues and Experimental Results for Acoustic Echo Cancellation," *Signal Processing*, vol. 61, no. 3, pp. 213–223, September 1997.

Moritz Harteneck and José Páez-Borrallo, "An Oversampled Subband
33] Adaptive Filter without Cross Adaptive Filters," *Signal Processing*, vol. 64, no. 1, pp. 93–101, January 1998.

Qing-Guang Liu, Benoit Champagne, and Dominic Ho, "Simple Design of
34] Oversampled Uniform DFT Filter Banks with Applications to Subband Acoustic Echo Cancellation," *Signal Processing*, vol. 80, no. 5, pp. 831–847, May 2000.

Zhang Li-Zhi, "A Novel Filter Bank Design Method," *IEEE Transaction on*
35] *Signal Processing*, vol. 1, no. 5, pp. 43-43, May 2009.

Simon Werner and José Apolinário, "Low-complexity constrained affine-
36] projection algorithms," *IEEE Transactions on Signal Processing, 2005*, vol. 53, no. 12, pp. 4545 - 4555, November 2005.

Reza Arablouei, "Affine Projection Algorithm with Selective Projections,"
37] *Signal Processing*, vol. 92, no. 9, pp. 2253–2263, September 2012.

Pablo Pérez's, Alberto Gonzalez, and Pedro Zuccarello, "Oversampled
38] nonuniform filter banks using quadratic optimisation and transition filters," *Electronics Letters*, vol. 43, no. 43, pp. 594–595, May 2007.

Xuemei Xie and Shi Guangming, "Method for Signal Decomposition and
39] Denoising Based on Nonuniform Cosine-Modulated Filter Banks," *Progress in Natural Science*, vol. 18, no. 10, pp. 1293-1298, October 2008.

Lalitha Ramya, "Alias Free Subband Adaptive Filtering," in *Recent
40] Advances in Space Technology Services and Climate Change*, Chennai, 2010, pp. 119–124.

Yung-Sze Choi, Se-Kwon Kim, and William Song, "Noise-Robust
41] Normalised Subband Adaptive Filtering," *Electronics Letters*, vol. 48, no. 8, pp. 432-434, April 2012.

Moon-Kyu Song, Seong-Eun Kim, Young-Seok Choi, and Woo-Jin Song,
42] "Selective Normalized Subband Adaptive Filter with Subband Extension," *IEEE Transactions on Circuits and Systems II*, vol. 60, no. 2, pp. 101 - 105, February 2013.

Yoshinori Ichikawa and Toshihiro Furukawa, "A New Approach for Non-
43] Uniform Subband Adaptive Filtering," in *IEEE International Symposium on Circuits and Systems*, Kobe, 2009, pp. 2271 - 2274.

Nguyen Thi and Christian Jutten, "Blind Source Separation for Convolutive
44] Mixtures," *Signal Processing*, vol. 45, no. 2, pp. 209-229, August 1995.

Sergio Cruces and Castedo Luis, "Stability Analysis of Adaptive

45] Algorithms for Blind Source Separation of Convolutive Mixtures," *Signal Processing*, vol. 28, no. 3, pp. 265-275, November 1999.

Serviere Christine, "Blind Source Separation of Convolutive Mixtures," in
46] *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, Saint-Martin, 1996, pp. 316 – 319.

Alexei Gorokhov and Philippe Loubaton, "Subspace-Based Techniques for
47] Blind separation of Convolutive Mixtures with Temporally Correlated Sources," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 9, pp. 813–820, September 1997.

Christian Jutten, Philippe Loubaton, and Mansour Ali, "Adaptive Subspace
48] Algorithm for Blind Separation of Independent Sources in Convolutive Mixture," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 583-586, February 2000.

Shoko Araki, Shoji Makino, Tsuyoki Nishikawa, and Hiroshi Saruwatari,
49] "The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutive Mixtures of Speech," in *IEEE Transactions on Speech and Audio Processing*, vol. 11, Salt Lake, March 2003, pp. 109-116.

Shoko Araki, Shoji Makino , Robert Aichner, Tsuyoki Nishikawa, and
50] Hiroshi Saruwatari , "Subband Based Blind Source Separation for Convolutive Mixtures of Speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 11, Utah, 2003, pp. 509-512.

Hicham Ghennioui, El Mostafa Fadaili, Nadège Thirion-Moreau, Abdellah
51] Adib, and Eric Moreau, "A Nonunitary Joint Block Diagonalization Algorithm for Blind Separation of Convolutive Mixtures of Sources," *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 860-863, October 2007.

Tiemin Mei, Alfred Mertins, and Fuliang Yin, "Blind Source Separation for
52] Convolutive Mixtures Based on the Joint Diagonalization of Power Spectral

Density Matrices," *Signal Processing*, vol. 88, no. 8, pp. 1990-2007, October 2008.

53] Herbert Buchner, Robert Aichner, and Walter Kellermann, "A Real-Time Blind Source Separation Scheme and its Application to Reverberant and Noisy Acoustic Environments," *Elsevier-Signal Processing*, vol. 86, no. 6, pp. 1260 - 1277, June 2009.

54] Xiaoming Zhu, Keshab Parhi, and Warren Warwick, "Blind Source Separation with Low Frequency Compensation for Convolutive Mixtures," in *The Forty-Third Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, 2009, pp. 1135-1139.

55] Hicham Ghennioui, Nadège Thirion-Moreau, and Eric Moreau, "Gradient-Based Joint Block Diagonalization Algorithms: Application to Blind Separation of FIR Convolutive Mixtures," *Signal Processing*, vol. 90, no. 6, pp. 1836-1849, June 2010.

56] Tariqullah Jan, DeLiang Wang, and Wenwu Wang, "A Multistage Approach to Blind Separation of Convolutive Speech Mixtures," *Speech Communication*, vol. 53, no. 4, pp. 524-539, April 2011.

57] Saito Shinya , Oishi Kunio , and Furukawa Toshihiro , "An Approach to Convolutive Backward-Model Blind Source Separation Based on Joint Diagonalization," in *The 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, 2012, pp. 579-583.

58] Shi Xizhi, "Blind Separation of Convolutive Mixtures of Speech Sources: Exploiting Local Sparsity," in *International Conference on Acoustics, Speech and Signal Processing*, Vancouver, 2013, pp. 4315-4319.

59] Shahab Minhas and Patrick Gaydecki, "A Hybrid Algorithm for Blind Source Separation of A Convolutive Mixture of Three Speech Sources," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1-15, December 2014.

Zhitang Chen and Laiwan Chan, "New Approaches for Solving Problems of
60] Convolved Mixtures," in *The International Joint Conference on Neural Networks*, San Jose, 2011, pp. 911-918.

Bahador Makki and Delaram Jarchi, "A Geometrically Constrained
61] Multimodal Time Domain Approach for Convolutive Blind Source Separation," in *19th European Signal Processing Conference*, Barcelona, 2011, pp. 1864 - 1868.

Roland Badeau, "Multichannel High-Resolution NMF for Modeling
62] Convolutive Mixtures of Non-Stationary Signals in the Time-Frequency Domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1670 - 1680, November 2014.

Toshihisa Tanaka, "A Direct Design of Oversampled Perfect
63] Reconstruction FIR Filter Banks," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 3011 - 3022, July 2006.

Shaeen Kalathil and Elizabeth Elias, "Efficient Design Of Non-Uniform
64] Cosine Modulated Filter Banks for Digital Hearing Aids," *International Journal of Electronics and Communications*, vol. 69, no. 9, pp. 1314–1320, May 2015.

Henning Reetz and Allard Jongman, *Phonetics: Transcription, Production,*
65] *Acoustics, and Perception*, 1st ed., Wiley-Blackwell, Ed. Chicester , United Kingdom: John Wiley and Sons Ltd, 2011.

William O'Grady, Francis Kata Dobrovolsky, and Francis Katamba,
66] *Contemporary Linguistics : an Introduction*, 1st ed., Pearson Education Limited, Ed. Harlow, United Kingdom: Longman, 2001.

John Clark, Colin Yallop, and Janet Fletcher, *An Introduction to Phonetics*
67] *and Phonology*, 3rd ed., Wiley-Blackwell, Ed. New York, United States: Wiley, 2007.

Joseph Desmond and Clare Fletcher, *Sounds English : a Pronunciation*
68] *Practice Book*, 1st ed. Sydney, Australia: Longman, 1989.

Peters Brady, Tamke Martin, Nielsen Stig Anton, Andersen Vestbjerg, and
69] Haase Mathias, "Responsive Acoustic Surfaces: Computing Sonic Effects," in *Respecting Fragile Places 29th eCAADe Conference*, Slovenia, 2011, pp. 819-828.

ISO 3382-1. (2009, June) International Organization (ISO) for
70] Standardization. [Online]. https://www.iso.org/standard/40979.html

Michael Vorländer , *Auralization: Fundamentals of Acoustics, Modelling,*
71] *Simulation, Algorithms and Acoustic Virtual Reality*, 1st ed. Aachen, Germany: Springer, 2007.

Charles Tisseyre, Hueber Rouard, and Courne Moulinier, "Intelligibility in
72] Various Rooms: Comparing its Assessment by (RA) STI Measurement with A Direct Measurement. Applied Acoustics," *Elsevier-Applied Acoustics*, vol. 53, no. 13, pp. 179-191, March 1998.

Tammo Houtgast and Herman Steeneken, "A Review of The MTF Concept
73] in Room Acoustics And its Use for Estimating Speech Intelligibility in Auditoria," *The Journal of the Acoustical Society of America*, vol. 77, no. 3, pp. 1069–1077, March 1985.

Beranek Leo, *Acoustics*, 1st ed. New York, United States: McGraw-Hill,
74] 1954.

Eckard Mommertz, *Acoustics and Sound Insulation : Principles, Planning,*
75] *Examples*, 1st ed. Base, Switzerland: Birkhäuser, 2009.

Emmanuel Vincenta, Shoko Araki, Fabian Theis, Guido Nolte, and Pau
76] Bofill, "The Signal Separation Evaluation Campaign (2007–2010): Achievements and Remaining Challenges," *Elsevier B.V*, vol. 92, no. 8, pp. 1928-1936, August 2012.

Emmanuel Vincent. (2007, January) Institute of Research in Computer
77] Science and Random Systems. [Online]. http://sassec.gforge.inria.fr/

German Cardoso and Antoine Souloumiac, "Blind Beamforming for Non-

78] Gaussian Signals," *IEE Proceedings on Radar and Signal Processing*, vol. 140, no. 6, pp. 362 - 370, December 1993.

Adel Belouchrani and Karim Abed-Meraim, "A Blind Source Separation 79] Technique Using Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434 - 444, May 1997.

Cott Douglas, Malay Gupta, Hiroshi Sawada, and Shoji Makino, "Spatio- 80] Temporal FastICA Algorithms for the Blind Separation of Convolutive Mixtures," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 5, pp. 1511 - 1520, July 2007.

Ibrahim Missaoui and Zied Lachiri, "Blind speech separation based on 81] undecimated wavelet packet perceptual," *International Journal of Computer Science Issues*, vol. 8, no. 3, pp. 1694-0814, May 2011.

Joseph Lacoume and Pablo Ruiz, "Sources Identification: a Solution Based 82] on Cumulants," in *Proceedings of the 4th ASSP Workshop on Spectral Estimator and Modeling*, Lisbon, 1988, pp. 199-203.

John John Daugman, "Entropy Reduction and Decorrelation in Visual 83] Coding by Oriented Neural Receptive Fields," *IEEE Transaction on Biomedical Engineering*, vol. 36, no. 1, pp. 107-114, January 1989.

Jean-Louis Lacoume, "Separation of Independent Sources from Correlated 84] Inputs," *IEEE Transaction on Signal Processing*, vol. 40, no. 12, pp. 3074-3078, December 1992.

Jay Jun, "Wide Band Blind Identification and Separation of Sources," in 85] *Proceedings of EUSIPCO*, Triestle, Italy, 1996, pp. 2077-2080.

Véronique Capdevielle, Christine Serviere, and Jean-Louis Lacoume, 86] "Blind Separation of Wideband Sources : Application to Rotating Machine Signals," in *Proceedings of Eusipco*, Triestle, Italy, 1996, pp. 2085-2088.

Christine Serviere and Veronique Capdevielle, "Blind Adaptive Separation

87] of Wide-Band Sources," in *Proceedings of ICASSP*, Atlanta, 1996, p. 2698—2701.

Hagai Attias and Christoph Schreiner, "Blind Source Separation and
88] Deconvolution : The Dynamic Component Algorithm," *Neural computation*, vol. 10, no. 1, pp. 1373-1424, April 1998.

Ivan Bradaric and Athina Petropulu, "On Resolving the Column
89] Permutation Ambiguity in the Estimates of MIMO System Response," in *Proceedings of ICA*, Aussois, 1999, pp. 11-15.

Lucas Parra and Clay Spence, "Convolutive Blind Separation of Non-
90] Stationary Sources," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 320-327, May 2000.

Lang Tong and Ruey-wen Liu, "Indeterminacy and Identifiability of Blind
91] Identification," *IEEE Transaction on Circuits Systems*, vol. 38, no. 5, pp. 499-509, May 1991.

Hsiao-Chun Wu and Jose Principe, "Simultaneous Diagonalization in the
92] Frequency Domain (SDIF) for Source Separation," in *34th Annual Conference on Information Sciences and Systems*, *CISS*, New Jersey, 2000, pp. 245-250.

Kiyotoshi Matsuoka, "Minimal Distortion Principle for Blind Source
93] Separation," in *Proceedings of the 41st SICE Annual Conference*, Osaka, 2002, pp. 2138 - 2143.

ITU-Recommendation P.800 - Methods for Subjective Determination of
94] Transmission Quality, 1996.

Nikolaos Mitianoudis and Michael Davies, "Audio Source Separation of
95] Convolutive Mixtures," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 489 - 497, September 2003.

Bao-Quan Liu and Da-Zheng Feng, "Adaptive Improved Natural Gradient
96] Algorithm for Blind Source Separation," *MIT Press*, vol. 21, no. 3, pp. 872 - 889,

January 2009.

[97] Yuu-Seng Lau, Zahir Hussian , and Richard Harris, "Performance of Adaptive Filtering Algorithms: A Comparative Study," in *Australian Telecommunications, Networks and Applications Conference (ATNAC)*, Melbourne, 2003, pp. 196–199.

[98] Sudha Siddappaji, "Performance Analysis of New Time Varying LMS (NTVLMS) Adaptive Filtering Algorithm in Noise Cancellation System for Speech Enhancement," in *The Fourth World Congress on Information and Communication Technologies (WICT)*, Bandar Hilir, 2014, pp. 224 - 228.

[99] Kong-Aik Lee, *Subband Adaptive Filtering Theory and Implementation*, 1st ed. West Sussex, UK: John Wiley & Sons, 2009.

[100] Ljiljana Milic, *Multirate Filtering for Dogital Signal Processing: MATLAB Applications*, 1st ed. New York, United States: Information Science Reference, 2009.

[101] Eftychios Papoulis and Tania Stathaki, "Extension of Generalised Subband Decomposition-Based Adaptive FIR Structure," *Electronics letters*, vol. 39, no. 15, pp. 1157 - 1158, May 2003.

[102] Brown Jeffrey, "Generalized Sampling and Perfect Reconstruction Problem for Maximally Decimated Filter Banks," in *International Conference on Acoustics, Speech, and Signal Processing, 1989. ICASSP-89*, Glasgow, 1989, pp. 1195 - 1198.

[103] Michael McCloud and Delores Etter, "Subband Adaptive Filtering with Time-Varying Nonuniform Filter Banks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Munich, 2010, pp. 1953 - 1956.

[104] Salah Al-Din Ibrahim Badran, Samad Ahmadi, Pooneh Bagheri Zadeh, and Ismail Shahin, "A Novel Tap Selection Design for Filters in Unequal-Passbands Scheme," in *The Eleventh International Conference on Digital Telecommunications (ICDT 2016)*, Lisbon, 2016, pp. 22-26.

Thomas Schlechter and Claire Galand, "Perfect Channel Splitting by Use of
105] Interpolation, Decimation, Tree Decomposition Techniques," in *International Conference on Information Sciences and Systems*, Patras, 1972, pp. 443–446.

Mark Smith and Tom Barnwell, "Filters for Distortion-Free Two-Band
106] Multirate Filter Banks," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 3, pp. 626–630, June 1985.

Mark Smith and Tom Barnwell, "A Procedure for Designing Exact
107] Reconstruction Filter Banks for Tree-Structured Subband Coders," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Diego, 1984, pp. 421 - 424.

Prashant Vaidyanathan, *Multirate Systems and Filter Banks*, 1st ed. New
108] Jersey, United States: Prentice Hall, 1993.

Saeed Vaseghi, *Multimedia Signal Processing: Theory and Applications in*
109] *Speech, Music, and Communications*, 1st ed. West Sussex, UK: John Wiley & Sons, 2007.

Norbert Fliege, *Multirate Digital Signal Processing: Multirate Systems -*
110] *Filter Banks - Wavelets*, 1st ed. Toronto, Canada: John Wiley & Sons, 1999.

Gordana Jovanovic-Dolecek, *Multirate Systems: Design and Applications:*
111] *Design and Applications*, 1st ed. London, UK: Idea Group Publishing, 2002.

Prashant Vaidyanathan, "Theory and Design of M-channel Maximally
112] Decimated Quadrature Mirror Filters with Arbitrary M, Having Perfect Reconstruction Property," *IEEE Trans.on Acoust., Speech, and Signal Processing*, vol. 35, no. 4, pp. 476–492, April 1987.

Prashant Vaidyanathan, "Quadrature Mirror Filter Banks, M-band
113] Extensions and Perfect-Reconstruction Techniques," *IEEE ASSP Magazine*, vol. 4, no. 3, pp. 4-20, July 1987.

Ramamurthi Koilpillai and Prashant Vaidyanathan, "Cosine-Modulated

114] Filter Banks Satisfying Perfect Reconstruction," *IEEE Transaction on Signal Processing*, vol. 40, no. 1, pp. 770 –783, August 1992.

115] Gerhard Doblinger, "A Fast Design Method for Perfect-Reconstruction Uniform Cosine-Modulated Filter Banks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6693 - 6697, May 2012.

116] Bingo Ling, Charlotte Ho, Kok Teo, and Wan-Chi Siu, "Optimal Design of Cosine Modulated Nonuniform Linear Phase FIR Filter Bank," *IEEE Transactions on Signal Processing*, vol. 62, no. 10, pp. 2517 - 2530, March 2014.

117] Truong Nguyen, "Near-Perfect-reconstruction Pseudo-QMF Banks," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 65 - 76, January 1994.

118] Shaheen and G. Ankita , "Review Article on Designing of Pseudo Quadratue Mirror Filter," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 9, no. 3, pp. 2278-8735, September 2014.

119] Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks*, 2nd ed. London, UK: Wellesley-Cambridge Press, 1996.

120] Jörg Kliewer and Alfred Mertins, "Oversampled Cosine-Modulated Filter Banks with Arbitrary System delay," *IEEE Transaction. on Signal Processing*, vol. 46, no. 1, pp. 941–955, April 1998.

121] Jose Apolinario and Rogerio Alves, "Filtered Gradient Algorithms Applied to a Subband Adaptive Filter Structure," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake, 2001, pp. 3705 - 3708.

122] Efiychios Papoulis and Tania Stathaki, "A Transmultiplexer-Based Adaptive Filtering Structure Using a New Adaptation Scheme," in *IEEE International Symposium on Circuits and Systems*, Phoenix-Scottsdale, 2012, pp. III-53 - III-56.

Supriya Dhabal and Srabani Chakraborty, "An efficient Quadrature Mirror Filter design and its applications in audio signal processing," in *International*

123] *Conference on Communication and Industrial Application (ICCIA)*, Kolkata, 2011, pp. 1 - 4.

Kostas Kokkinakis and Philipos Loizou, "Subband-Based Blind Signal
124] Processing for Source Separation in Convolutive Mixtures of Speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Honolulu, 2007, pp. IV-917 - IV-920.

Bo Peng and Wei Liu, "An Improved Solution to the Subband Blind Source
125] Separation Permutation Problem Based on Optimized Filter Banks," in *th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Limassol, 2010, pp. 1 - 4.

Changzhi Wu and Kok Teo, "Design of Discrete Fourier Transform
126] Modulated Filter Bank With Sharp Transition Band," *IET Signal Processing*, vol. 5, no. 4, pp. 433 - 440, July 2011.

Yu Yonglin and Wang Gang, "Innovation Method for Independent
127] Component Analysis: A New Concept and Algorithm," in *WRI Global Congress on Intelligent System*, Xiamen, 2009, pp. 428 - 431.

Andrzej Cichocki and Rolf Unbehauen, "Robust Learning Algorithm for
128] Blind Separation Of Signals," *Electronics Letters*, vol. 30, no. 17, pp. 1386 - 1387, August 1994.

James Stone, "Independent Component Analysis:A Tutorial Introduction,"
129] in *Independent Component Analysis and Blind Source Separation*. London, UK: MIT Press, 2004, ch. 7, pp. 1 - 3.

Jagath Rajapakse, "Adaptive Blind Signal and Image Processing: Learning
130] Algorithms and Applications," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1580 - 1584, November 2003.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component*
131] *Analysis*, 2nd ed. New York, United States: Wiley, 2001.

Jacob Benesty and Yiteng Huang, *Adaptive Signal Processing: Applications*
132] *to Real-World Problems*, 1st ed. Murray Hill, United States: Springer, 2003.

Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation:*
133] *Independent Component Analysis and Applications*, 1st ed. Oxford, UK: Academic press, 2010.

Fuxiang Wang and Jun Zhang, "Mutual Information Minimization forBlind
134] Source Separation," in *International Conference on Computational Intelligence and Software Engineering*, Wuhan, 2009, pp. 1 - 4.

Fuliang Yin, Tiemin Mei, and Jun Wang, "Blind-Source Separation Based
135] on Decorrelation and Nonstationarity," *IEEE Transactions on Circuits and Systems*, vol. 54, no. 5, pp. 1150 - 1158, May 2007.

Miloje Radenkovic and Tamal Bose, "Blind Separation of MIMO Systems:
136] New Algorithms and Convergence Analysis," *IEEE Transactions on Circuits and Systems*, vol. 57, no. 7, pp. 1475 - 1488, July 2010.

Young-Gul Won and Soo-Young Lee, "Convolutive Blind Signal
137] Separation by Estimating Mixing Channels in Time Domain," *Electronics Letters*, vol. 44, no. 21, pp. 1277 - 1279, October 2008.

Tulay Adali and Peter Schreier, "Optimization and Estimation of Complex-
138] Valued Signals: Theory and applications in filtering and blind source separation," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 112 - 128, August 2014.

Herbert Buchner, Robert Aichner, and Walter Kellermann, "A
139] Generalization of a Class of Blind Source Separation Algorithms for Convolutive Mixtures," in *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, 2003, pp. 945–950.

Shi Xizhi, *Blind Signal Processing: Theory and Practice*, 1st ed. Shanghai,
140] China: Springer, 2011.

Francesco Nesta and Piergiorgio Svaizer, "Convolutive BSS of Short

141] Mixtures by ICA Recursively Regularized Across Frequencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 624 - 639, June 2011.

142] Shahram Hosseini and Yannick Deville, "Blind Separation of Parametric Nonlinear Mixtures of Possibly Autocorrelated and Non-Stationary Sources," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6521 - 6533, November 2014.

143] Bile Peng and Wenxin Liu, "Reducing Permutation Error in Subband-Based Convolutive Blind Separation," *IET Signal Processing*, vol. 6, no. 1, pp. 34-44, February 2012.

144] Julien Bourgeois and Klaus Linhard, "Frequency-Domain Multichannel Signal Enhancement: Minimum Variance vs. Minimum Correlation," in *8th International Conference on Signal Processing*, Vienna, 2005, pp. 389 - 392.

145] Christine Serviere and Dinh-Tuan Pham, "Permutation Correction in the Frequency Domain in Blind Separation of Speech Mixtures," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 1, pp. 1-16, April 2006.

146] Hao Yuan, Hisashi Kawai, and Toshiharu Horiuchi, "Reduction of Correlation Computation in the Permutation of the Frequency Domain ICA by Selecting DOAs Estimated in Subarrays," in *15th European Signal Processing Conference*, Poznan, 2007, pp. 418 - 422.

147] Wei Zhao , Yuehong Shen , and Pengcheng Xu , "A New Efficient Method for Permutation and Scaling Ambiguity of BSS Blocks," in *Fifth International Conference on Intelligent Control and Information Processing (ICICIP)*, Dalian, 2014, pp. 23 - 28.

148] Radoslaw Mazur and Alfred Mertins, "Using the Scaling Ambiguity for Filter Shortening in Convolutive Blind Source Separation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Taipei, 2009, pp.

1709 - 1712.

149] Hai-Huyen Dam, Antonio Cantoni, and Sven Nordholm, "Second-Order Blind Signal Separation for Convolutive Mixtures," *IEEE Signal Processing Letters*, vol. 15, no. 1, pp. 79 - 82, January 2008.

150] Emmanuel Vincent and Rémi Gribonval, "Performance Measurement in Blind Audio Source Separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462 - 1469, June 2006.

151] Rajeev Nongpiur and Dale Shpak, "Maximizing the Signal-to-Alias Ratio in Non-Uniform Filter Banks for Acoustic Echo Cancellation," *IEEE Transactions on Circuits and Systems*, vol. 59, no. 10, pp. 2315 - 2325, August 2012.

152] Campbell Douglas, Palomäki Kalle, and Brown Guy, "A fast and accurate “shoebox” room acoustics simulator," *Computer Information System*, vol. 9, no. 3, pp. 48–51, September 2005.

153] International Telecommunications Union (ITU). (2005, November) Recommendation P.862. [Online]. http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en

154] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *International Conference on Neural Networks for Computing*, Utah, 1986, pp. 206-211.

155] Remy Wenmaekers and Constant Hak, "On Measurements of Stage Acoustic Parameters: Time Interval Limits and various Source–Receiver Distance," *Ingenta Connect*, vol. 98, no. 5, pp. 776-789, October 2012.

156] Longbiao Wang, Kyohei Odani, Atsuhiko Kai, and Weifeng Li, "Speech recognition using blind source separation and dereverberation method for mixed sound of speech and music," in *IEEE- Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Kaohsiung, Taiwan, 2013, pp. 1 - 4.

Cheol-Ho Jeonga, Jonas Brunskog, and Finn Jacobsen, "Room Acoustic
157] Transition Time based on Reflection Overlap," *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 2733-2736, May 2010.

Murthi Sridharan, "Subband Adaptive Filtering: Oversampling Approach,"
158] *Signal Processing*, vol. 71, no. 1, pp. 101–104, November 1998.

André Kouwe, DeLiang Wang, and Guy Brown, "A Comparison of
159] Auditory and Blind Separation Techniques for Speech," *IEEE Transactions on Speech And Audio Processing*, vol. 9, no. 3, pp. 189-195, March 2001.

Frederic Berthommier and Seungjin Choi, "Evaluation of CASA And BSS
160] Models for Subband Cocktail Party Speech Recognition," in *Proceedings of ICSP' 01*, Daejeon, 2001, pp. 301-306.

Jeanny Herault and Christian Jutten, "Space or Time Adaptive Signal
161] Processing by Neural Network Models," in *Symposium on Signal and Image Processing*, Nice, France, 1985, pp. 1017-1022.

Shaoming Li, Bo Yang, and Jiayan Zhang, "BSS Algorithm Based on Fully
162] Connected Recurrent Neural Network and the Application in Separation of Speech Signals," in *Spring Congress on Engineering and Technology (S-CET)*, Xian, 2012, pp. 1 - 3.

Yannick Deville, "Towards Industrial Applications of Blind Source
163] Separation and Independent Component Analysis," in *First International Conference on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 19-24.

Yina Guo and Ganesh Naik, "Single Channel Blind Source Separation
164] Based Local Mean Decomposition for Biomedical Applications," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, 2013, pp. 6812 - 6815.

Christian Jutten and Massoud Babaie-Zadeh, "Source Separation:

165] Principles, Current Advances and Applications," German-French Institute for Automation and Robotic Annual Meeting, Nancy, France, 2006.

166] Sergio Cerutti, Carlo Marchesi, and Luca Mesin, *Blind Source Separation: Application to Biomedical Signals*, 1st ed. NewYork, United States: Wiley-IEEE Press, 2013.

167] Gaoming Huang and Luxi Yang, "A Radar Anti-Jamming Technology Based on Blind Source Separation," in *7th International Conference on Signal Processing*, Shanghai, 2004, pp. 2021 - 2024.

168] Johnson James and Rocafort Jorge, *Architectural acoustics : Principles and Design*, 1st ed. Sydney, Australia: Prentice Hall, 1999.

169] Malcolm Crocker, *Handbook of Acoustics*, 1st ed. Auburn, United States: Wiley, 1998.

170] Erwin Meyer, *Physical and Applied Acoustics*, 1st ed. Massachusetts, United States: Academic Press, 1972.

171] Yiteng Huang, Jacob Benesty, and Jingdong Chen, *Acoustic MIMO Signal Processing*, 1st ed. New York, United States: Springer, 2006.

172] Kinsler Lawrence, Frey Austin, Coppens Alan, and Sanders James, *Fundamentals of Acoustics*, 1st ed. New York, United States: Wiley, 1982.

173] Webstore-International Electrotechnical Commision. (2011, June) IEC 60268-16 -Sound System equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index. [Online]. https://webstore.iec.ch/publication/1214

174] Bruel Kjaer, Dirac Room Acoustics Software, 2003, Sound & Vibration Measurement A/S.

175] Ulf Lindgren, "Source Separation using a Criterion Based on Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 1837 -

1850, April 1998.

176] Mehrez Souden, Shoko Araki, and Keisuke Kinoshita, "A Multichannel MMSE-Based Framework for Speech Source Separation and Noise Reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1913 - 1928, May 2013.

177] Ali Mansour, Mitsuru Kawamoto, and Noburo Ohnishi, "Blind Separation for Instantaneous Mixture of Speech Signals: Algorithms and Performances," in *TENCON 2000. Proceedings*, Kuala Lumpur, 2000, pp. 26 -32.

178] Yiling Xu, Le Yang, and Qicong Peng, "Innovation Enhanced Infomax for Separating Linear Instantaneous Mixtures," in *International Conference on Communications, Circuits and Systems*, Shanghaii, 2004, pp. 1018 - 1021.

179] Shun-Ichi Amari and Andrzej Cichocki, "Adaptive Blind Signal Processing-Neural Network Approaches," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2026 - 2048, May 1998.

180] Lan-Da Van, Di-You Wu, and Chien-Shiun Chen, "Energy-Efficient FastICA Implementation for Biomedical Signal Separation," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1809 - 1822, April 2011.

181] Lang Tong and Vic Soon, "Amuse : a New Blind Identification Algorithm," *IEEE International Symposium on Circuits and Systems*, vol. 3, no. 1, pp. 1784 - 1787, May 1990.

182] Andreas Ziehe and Klaus-Robert Müller, "An Efficient Algorithm for Blind Separation Using Time Structure," in *The 8th International Conference on Artificial Neural Networks, ICANN'98*, Berlin, 1998, pp. 675 - 680.

183] Adel Belouchrani and Andrzej Cichocki, "Robust Whitening Procedure in Blind Source Separation Context," *IEEE, Electronics Letters*, vol. 36, no. 24, pp. 2050 - 2051, November 2000.

Intae Lee, Taesu Kim, and Te-Won Lee, "Complex Fastica : a Robust

184] Maximum Likelihood Approach of Mica for Convolutive Bss," in *6th International Conference, ICA 2006*, Charleston, 2006, pp. 62 - 632.

Taesu Kim, Hagai Attias, and Soo-Young Lee, "Blind Source Separation
185] Exploiting Higher-Order Frequency Dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70 - 79, January 2007.

Joe Qin and Ricardo Dunia, "Determining the Number of Principal
186] Components for Best Reconstruction," *Journal of Process Control*, vol. 10, no. 2, pp. 245 - 250, March 2000.

Moeness Amin, "Spatial Time-Frequency Distributions for Direction
187] Finding and Blind Source Separation," in *The SPIE Wavelet Conference*, Orlando, 1999, p. April.

Ehud Weinstein and Alan Oppenheim, "Multi-Channel Signal Separation
188] by Decorrelation," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 405-413, October 1993.

Noboru Murata, Shiro Ikeda, and Andreas Ziehe, "An Approach to Blind
189] Source Separation Based on Temporal Structure of Speech Signals," *Neurocomputing*, vol. 41, no. 4, pp. 98-102, October 1998.

Nguyen Thi and Christian Jutten, "Speech Enhancement: Analysis and
190] Comparison of Methods in Various Real Situations," *Signal Processing VI, Theories and Applications (EUSIPCO'92)*, vol. 6, no. 3, pp. 303-306, August 1992.

Hajipour Sardouie, Bagher Shamsollahi , and Albera Merle, "Denoising of
191] Ictal EEG Data Using Semi-Blind Source Separation Methods Based on Time-Frequency Priors," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 839 - 847, July 2015.

Kang Chun-yu Chun-yu, Fan Wen-Tao Wen-Tao, and Zhang Xin-hua, "A
192] Kind of Method for Direction of Arrival Estimation Based on Blind Source Separation Demixing Matrix," in *IEEE- 8th International Conference on Natural*,

Chongqing, 2012, pp. 134 - 137.

193] Usevitch Bryan and Orchard Michael, "Adaptive Filtering Using Filter Banks," *IEEE Transactions on Circuits and Systems II*, vol. 43, no. 3, pp. 255-265, March 1996.

194] Wallace Clement Sabine, *Collected Papers on Acoustics : with a New Introduction*, 1st ed. New York, United States: Dover Publications, 1964.

195] Heinrich Kuttruff, *Room Acoustics*, 5th ed. Florida, United States: CRC Press , 2014.

196] Jingang Zhong and Yu Huang, "Time-Frequency Representation Based on an Adaptive Short-Time Fourier Transform," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5118 - 5128, October 2010.